



دانشکده مهندسی کامپیوتر

عنوان کارآموزی سیستم‌های پیشنهاد دهنده با استفاده از یادگیری ماشین

محل کارآموزی

آزمایشگاه رایانش نرم و سیستم‌های چند عامله
واقع در دانشگاه علم و صنعت ایران دانشکده مهندسی کامپیوتر
تحت نظر آقای دکتر مزینی

نام دانشجو

محمد محمدی

نام استاد کارآموزی:

آقای دکتر جاهد مطلق

خرداد ماه ۱۴۰۰

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

تأییدیه‌ی صحت و اصالت نتایج

بسمه تعالی

اینجانب محمد محمدی به شماره دانشجویی ۹۵۵۲۱۴۱۴ دانشجوی رشته مهندسی کامپیوتر مقطع تحصیلی کارشناسی تأیید می‌نمایم که کلیه‌ی مطالب مندرج در این گزارش حاصل ۳۰۰ ساعت حضور و کار اینجانب در شرکت/کارخانه آزمایشگاه رایانش نرم و سیستم‌های چند عامله و بدون هرگونه دخل و تصرف است و موارد نسخه‌برداری شده از آثار دیگران را با ذکر کامل مشخصات منبع ذکر کرده‌ام. در صورت اثبات خلاف مندرجات فوق، به تشخیص دانشگاه مطابق با ضوابط و مقررات حاکم آموزشی، پژوهشی و انضباطی با اینجانب رفتار خواهد شد و حق هرگونه اعتراض درخصوص احقاق حقوق مکتسب و تشخیص و تعیین تخلف و مجازات را از خویش سلب می‌نمایم.

نام و نام خانوادگی: محمد محمدی

امضا و تاریخ: ۲۷ / ۳ / ۱۴۰۰

تشکر و قدردانی:

با تشکر و قدردانی از آقای دکتر مزینی که امکان حضور من در آزمایشگاه را فراهم آوردند.

چکیده

حوزه کارآموزی دانشگاه علم و صنعت بود. واحد استقرار آزمایشگاه سیستم های چند عامله و رایانش نرم زیر نظر دکتر مزینی در طبقه اول دانشکده مهندسی کامپیوتر بود.

در این آزمایشگاه به تحقیق در خصوص سیستم های چندعامله و یادگیری تقویتی و ... پرداخته میشود. در این دوره از کارآموزی به یادگیری اصول یادگیری ماشین و مقدمه ای از پژوهش پرداخته شد. همچنین به بررسی با جزئیات انواع سیستم های پیشنهاد دهنده و نحوه ی پیاده سازی و عملکرد آنها با استفاده از یادگیری ماشین پرداخته شد. در ابتدا به خواندن مقالات و کدهای موجود در حوزه یادگیری ماشین پرداخته شد سپس با استفاده از کارهای موجود به بررسی انواع سیستم های پیشنهاد دهنده و مقایسه پارامتری آنها پرداخته شد.

واژه های کلیدی: هوش مصنوعی ، یادگیری ماشین ، Recommender System

7	فصل ۱: معرفی حوزه کارآموزی
8	۱-۱- مقدمه.....
10	فصل ۲: مشروح فعالیت های انجام شده در محل استقرار
11	۲-۱- مقدمه.....
11	۲-۲- معرفی واحد محل استقرار کارآموز.....
11	۲-۳- شرح پروژه ها و فعالیت های انجام شده توسط کارآموز.....
13	۲-۴- نتیجه گیری.....
15	مراجع

فصل ۱: معرفی حوزه کارآموزی

۱-۱- مقدمه

دانشگاه علم و صنعت ایران در زمینی به مساحت ۴۲۰ هزار متر مربع در شمال شرقی شهر تهران واقع شده است. هسته اولیه دانشگاه در سال ۱۳۰۸ با هدف ایجاد زمینه‌های لازم برای تحصیلات دوره عالی مهندسی در کشور، با نام هنرسرای عالی در خیابان ۳۰ تیر (قوام السلطنه سابق) تأسیس شد. در سال ۱۳۳۶ هنرسرای عالی با تغییراتی به محل فعلی دانشگاه صنعتی امیرکبیر (پلی تکنیک سابق) انتقال یافت و به عنوان انستیتو تکنولوژیک تهران هنرسرای عالی فنی، فعالیت خود را ادامه داد. از سال ۱۳۴۱ بار دیگر این مؤسسه دستخوش تحولاتی گردید و به مکان جدید در نارمک (محل کنونی دانشگاه علم و صنعت ایران) انتقال یافت. از این زمان به بعد هدف اصلی در نظر گرفته شده برای آن تربیت دبیر فنی برای تأمین کادر هنرستان‌های صنعتی بود. در سال ۱۳۵۱ هنرسرای عالی به دانشکده علم و صنعت ایران تغییر نام یافت و در نهایت پیش از پیروزی انقلاب اسلامی در سال ۱۳۵۷ و با توجه به گسترش و تغییرات کیفی کارگاه‌های علمی و آموزشی، از سوی وزارت آموزش عالی به دانشگاه علم و صنعت ایران تبدیل شد.

دانشگاه علم و صنعت ایران در بیش از ۳۲ گرایش در مقطع کارشناسی، ۶۶ گرایش در مقطع کارشناسی ارشد و ۳۲ گرایش در مقطع دکترا دانشجوی می‌پذیرد. تعداد اعضای هیات علمی ۳۷۴ نفرند که به تفکیک عبارتند از: ۲۵ استاد، ۸۴ دانشیار، ۲۰۲ استادیار، ۶۲ مربی و یک مربی آموزشیار. این دانشگاه دارای امکانات گسترده‌ای از قبیل: دو کتابخانه مرکزی و ۱۷ کتابخانه اقماری مستقر در دانشکده‌ها و مراکز تحقیقاتی، مرکز انتشارات، مرکز کامپیوتر، مرکز فناوری اطلاعات، کارگاه‌های صنعتی، آزمایشگاه‌های متعدد و مجهز آموزشی و پژوهشی و تخصصی و ۱۰ مرکز تحقیقاتی می‌باشد. عناوین این مراکز عبارتند از: مرکز تحقیقات سیمان، پژوهشکده سبز، پژوهشکده الکترونیک، مرکز تحقیقات مکانیک خودرو، موسسه کامپوزیت، مرکز تحقیقات مواد نسوز، آلومینیوم و ریخته‌گری، مرکز تحقیقات معماری و شهرسازی، مرکز تحقیقات فناوری اطلاعات (IT)، مرکز تحقیقات فناوری اطلاعات و ارتباطات (ICT)، مرکز تحقیقات ژئوتکنیک و مرکز آموزش الکترونیکی (مجازی). مرکز اخیر از بهمن ۱۳۸۳ در سه رشته مهندسی کامپیوتر - نرم افزار، مهندسی فناوری اطلاعات و علوم کامپیوتر دانشجوی پذیرفته است.



۱۳ دانشکده دانشگاه علم و صنعت ایران:

- دانشکده مهندسی برق
- دانشکده مهندسی راه‌آهن
- دانشکده مهندسی خودرو
- دانشکده ریاضی
- دانشکده مهندسی شیمی
- دانشکده شیمی
- دانشکده مهندسی صنایع
- دانشکده مهندسی عمران
- دانشکده مهندسی مواد و متالورژی

- دانشکده مهندسی مکانیک
- دانشکده مهندسی کامپیوتر
- دانشکده مهندسی معماری و شهرسازی
- دانشکده فیزیک

سطح آموزشی بالا از یک سو، و داشتن امکانات آزمایشگاهی و کارگاهی پیشرفته از سویی دیگر به خصوص در رشته‌های مهندسی عمران، مواد و متالورژی و مکانیک باعث شده تا وجهه‌ی صنعتی این دانشگاه پررنگ‌تر شود و دانش‌آموختگان آن بتوانند به سرعت جذب واحدهای صنعتی ایران شوند. همچنین اخیراً پروژه‌های نوینی همچون طراحی ماهواره ملی نوید و ظفر علم و صنعت و پروژه سپر ملی دفاع سایبری در مرکز ملی دفاع سایبری کشور همچنین طراحی موتور هواپیماهای جت و طراحی پلتفرم ملی خودرو کلاس B از طرح‌ها و برنامه‌های ملی و ویژه این دانشگاه محسوب می‌شوند. [۱]

دانشکده مهندسی کامپیوتر در سال ۱۳۶۸ بطور رسمی شروع به کار نمود. این دانشکده به ترتیب در مقطع کارشناسی و سپس در مقاطع کارشناسی ارشد و دکتری کامپیوتر دانشجو پذیرفت و علی‌رغم سابقه نه چندان طولانی توانسته است در سطح کشور به عنوان یکی از مراکز علمی فعال در حوزه کامپیوتر و فناوری اطلاعات تبدیل شود. [۲]

آزمایشگاه‌های تحقیقاتی :

- آزمایشگاه تحقیقاتی پردازش تصویر
- آزمایشگاه تحقیقاتی پردازش صدا و گفتار
- آزمایشگاه تحقیقاتی سیستمهای پیچیده (آشوب)
- آزمایشگاه تحقیقاتی اتوماسیون هوشمند
- آزمایشگاه تحقیقاتی پردازش موازی و همروند
- آزمایشگاه تحقیقاتی مدل‌های شناختی محاسبه‌ای
- آزمایشگاه رایانش اعتماد پذیر
- آزمایشگاه تحقیقاتی رایانش نرم و سیستمهای چند عامله

فصل ۲: مشروح فعالیت های انجام شده در محل استقرار

۲-۱- مقدمه

در این قسمت به بررسی دقیق کارها انجام شده در محل کار آموزشی پرداخته خواهد شد.

۲-۲- معرفی واحد محل استقرار کارآموز

محل کارآموزی آزمایشگاه سیستم های چند عامله و رایانش نرم زیر نظر دکتر مزینی بود. در این حوزه از کامپیوتر که به رایانش نرم شناخته میشود به حل مسائلی که به شکل فرمال قابل حل نیستند یا قابل مدل سازی هستند اما مدل پیچیده میباشد از طریق روش های دیگر مانند منطق فازی و دیپ لرنینگ و ... پرداخته میشود.

۲-۳- شرح پروژهها و فعالیت های انجام شده توسط کارآموز

در ابتدا موضوع پروژه تحقیقاتی تعیین شد که سیستم های پیشنهاد دهنده انتخاب شد سپس با توجه به فراگیر شدن و گرفتن نتایج خوب از روش های یادگیری عمیق برای حل این مساله نیز از یادگیری عمیق استفاده شد.

خیلی از کدها و روش های موجود مشکل با ورژن های کنونی فریم ورک های یادگیری ماشین داشتند که موجب صرف وقت بر روی نصب فریمورک های مختلف شد که با توجه به حجم زیاد این فریمورک ها یکی از مشکلات اصلی در این دوره بود. همچنین بسیاری از کدهای موجود دارای باگ بودند.

در گام بعدی ابتدا به سراغ [https://www.kaggle.com/gspmoreira/recommender-systems-in-](https://www.kaggle.com/gspmoreira/recommender-systems-in-python-۱۰۱#Loading-data-CI&T-Deskdrop-dataset)

[python-۱۰۱#Loading-data-CI&T-Deskdrop-dataset](https://www.kaggle.com/gspmoreira/recommender-systems-in-python-۱۰۱#Loading-data-CI&T-Deskdrop-dataset) رفتم چرا که همانند مقدمه ای کامل و جامع بر سیستم های پیشنهاد دهنده بود و به توضیح و پیاده سازی انواع آن ها با استفاده از روش های یادگیری ماشین و یادگیری عمیق پرداخته بود.

هدف اصلی سیستم های پیشنهاد دهنده پیشنهاد دادن موارد مرتبط به کاربر براساس اولویت های او می باشد. که این اولویت و ارتباط در انتخاب ها نسبی بوده و اصولا براساس رفتار قبلی کاربر صورت می گیرد. سه دسته ای اصلی از سیستم های پیشنهاد دهنده به عبارت زیرند که به توضیح جامع هر کدام خواهیم پرداخت:

- **فیلتر کردن مشارکتی (Collaborative Filtering):** در این روش پیشنهادها با جمع آوری اولویت ها و سلیق تعداد زیادی از کاربران ساخته می شوند. بدین صورت که اگر فرد الف در مجموعه ای از انتخاب ها اولویت های مشابه با فرد ب داشت احتمالاً در مورد یک انتخاب جدید نیز اولویت هایی از جنس اولویت های فرد ب دارد.

- **فیلتر کردن مبتنی بر محتوا (Content-Based Filtering):** در این روش فقط اطلاعات و محتوای مربوط به انتخاب هایی که کاربر در گذشته انجام داده است بررسی می شوند تا مدل دقیقی از اولویت های کاربر ساخته شود و پیشنهادها بدین صورت ساخته می شوند که مشابه با مواردی باشند که کاربر در گذشته انتخاب کرده است و بدین صورت سیستم تلاش می کند تا نزدیک ترین انتخاب به اولویت های کاربر را به او پیشنهاد دهد.

- **فیلتر کردن با روش ترکیبی (Hybrid Method):** در این روش از هر دو رویکرد قبلی بصورت ترکیبی و در کنار هم استفاده می شود و همانطور که در مقالات اخیر مشاهده شده است دقت بالاتری نسبت به دو روش قبلی از این روش بدست می آید. همچنین با این روش مشکلات رایجی از جمله مشکل شروع سرد (Cold Start) و مشکل پراکندگی (Sparsity Problem) که در روش های قبلی وجود داشت نیز برطرف میگردد.

در ادامه پیاده سازی هر سه سیستم پیشنهاد دهنده را مطالعه و تست نمودم و نتایج حاصل را در زیر بیان می نمایم.

دیتاست (Dataset): در این پیاده سازی ها از دیتاست [Articles sharing and reading from CI&T](#) [DeskDrop](#) استفاده شده است که حاصل ۱۲ ماه log سیستم CI&T از اینتراکشن های بین ۷۳ هزار کاربر در به اشتراک گذاری ۳ هزار مقاله بین یکدیگر می باشد.

ارزیابی (Evaluation): در این پیاده سازی از یک روش Cross-Validation به نام HoldOut استفاده شده است بدین صورت که ۲۰ درصد از دیتای train جدا گذاشته می شود و برای ارزیابی از آن استفاده می شود.

رویکرد پایه (Baseline Approach): در این پیاده سازی از مدل محبوبیت (Popularity Model) یا همان مدل ترند به عنوان رویکرد پایه استفاده شده است که شخصی سازی شده نبوده و تنها در هر لحظه محبوب ترین انتخابی که کاربر تا به این لحظه انتخاب نکرده را به او پیشنهاد می دهد. این رویکرد بر پایه ی خرد جمعی (Wisdom of the Crowds) بوده و معمولاً پیشنهادات خوبی را می دهد ولی روش های بر پایه ی یادگیری ماشین باید از این روش بهتر عمل کنند و پیشنهاداتی شخصی سازی شده و بر اساس سلیق کاربر به او بدهند و نه فقط پیشنهادات محبوب را.

در اکثر تحقیقات و مقالات این زمینه رویکرد پایه همین مدل می باشد. در این پیاده سازی روی این دیتاست رویکرد پایه به دقت های ۲۴ و ۳۷ درصد به ترتیب در انتخاب ۵ و ۱۰ آیتم برای پیشنهاد رسید و تلاش بر این است که با رویکردهای مبتنی بر یادگیری ماشین به دقت های بالاتری برسیم.

روش فیلتر کردن مبتنی بر محتوا (Content-Based Filtering Model):

در رویکرد تمرکز بر روی توضیحات و ویژگی های آیتم هایی است که کاربر قبلاً انتخاب کرده و با بررسی آن ها و مقایسه آن با آیتم های جدید تلاش می شود که موارد مشابه به کاربر پیشنهاد دهد. با توجه به اینکه این روش فقط انتخاب های قبلی کاربر را بررسی می کند و بر اساس آن ها عمل می کند در برابر شروع سرد (Cold Start) مقاوم است.

برای ساختن پروفیلی از کاربر از روش میانگین وزن دار (Weighted Average) استفاده می شود و در میان مقالاتی که کاربر با آن ها سروکار داشته و بر اساس میزان اینترکشن کاربر میانگین وزن دار گرفته می شود. نتایج شخصی سازی شده ای که از این روش دریافت شد نشان دهنده ی دقت ۱۶ و ۵۲ درصدی به ترتیب در ۵ و ۱۰ انتخاب اول بود که نتیجه ای که می توان از ضعیف تر بودن برخی نتایج در این رویکرد نسبت به رویکرد محبوبیت گرفت این است که کاربران در انتخاب هایشان فقط یک محتوای نسبتاً ثابت را دنبال نمی کنند.

روش فیلتر کردن مشارکتی (Collaborative Filtering Model):

در پیاده سازی این روش دو استراتژی می توان پیش گرفت:

- **استراتژی مبتنی بر حافظه:** در این استراتژی از بررسی انتخاب های کاربران در گذشته استفاده می کنیم تا مشاهبت های بین کاربران را بیابیم (مدل کاربر محور) و یا انتخاب ها و آیتم ها را بررسی می کنیم تا مشاهبت های بین آن ها را بیابیم.
- **استراتژی مبتنی بر مدل:** در این استراتژی مدل هایی با استفاده از الگوریتم های یادگیری ماشین توسعه یافته می شود تا به کاربر پیشنهادهایی بدهند.

در اینجا از استراتژی مبتنی بر مدل و مدل SVD Matrix Factorization در پیاده سازی استفاده شده است و که نتایج پیشنهادهای تولید شده توسط این روش به دقت های بینظیر ۳۳ و ۴۶ درصد به ترتیب در ۵ و ۱۰ پیشنهاد اول اشاره دارد که دقتی به مراتب بالاتر از رویکرد پایه و رویکرد مبتنی بر محتوا می باشد.

فیلتر کردن با روش ترکیبی (Hybrid Method):

در این روش هر دو روش قبلی را با هم ترکیب می‌کنیم و به نتایج بهتری هم میرسیم. روش پیاده‌سازی بدین صورت است که از هر دو روش قبلی یک میانگین وزن‌دار میگیریم و پیشنهاد نهایی به صورت وزن‌داری حاصل هر دو روش قبلی خواهد بود. که در این پیاده‌سازی دقت روش فیلتر کردن مشارکتی بسیار بیشتر بود و به همین دلیل وزن‌ها در روش ترکیبی بصورت ۱ و ۱۰۰ به ترتیب برای روش مبتنی بر محتوا و روش مشارکتی خواهند بود. نتایج پیشنهادهای شخصی‌سازی شده‌ای که از این روش بدست آمد به دقت‌های ۳۴.۲ و ۴۷.۹ درصد به ترتیب در ۵ و ۱۰ پیشنهاد اول رسید که از هر دو روش قبلی بصورت جدا جدا بالاتر بود.

مقایسه مدل‌ها:

دقت‌های بدست آمده از هر ۳ روش در کنار رویکرد پایه بصورت زیر در آمد:

	recall@5	recall@10
modelName		
Content-Based	0.162874	0.261442
Popularity	0.241754	0.372923
Collaborative Filtering	0.334058	0.468167
Hybrid	0.342751	0.479673

۴-۲- نتیجه گیری

در این دوره و با بررسی این مقاله جامع و تحقیق در آن و تلاش برای اجرای آن و همچنین تحقیق - و توسعه در برخی از بخش‌های آن روش‌های اصلی در ساخت سیستم‌های پیشنهاد دهنده را بررسی کردیم و دریافتیم که یک روش ترکیبی می‌تواند در صورت اجرای صحیح عملکرد بهتری داشته باشد.

در ادامه، در صورتی که افراد دیگر در دوره‌های بعدی آزمایشگاه مایل به کار بر روی سیستم‌های

پیشنهاد دهنده باشند می‌توانند در صورت تمایل از موارد زیر ادامه دهند و نتایج فعلی را بهبود دهند:

- در این مقاله مسأله‌ی زمان مطرح نبود اما باید در نظر گرفت که در یک دنیای واقعی مسأله‌ی زمان مطرح است و باید فقط از میان انتخاب‌های موجود در لحظه به کاربر پیشنهاد داد.
- می‌توان اطلاعات بیشتری از کاربر را نیز در پیشنهادها دخیل کرد. اطلاعاتی از قبیل سلیقه کاربر در زمان حال (نه در کل زمان) و اطلاعات منطقه‌ای و جغرافیایی و اطلاعات دستگاه مورد اتصال کاربر (دسکتاپ، گوشی، تبلت و ...).

و همچنین مطالعه و بررسی مقالات گوناگونی که در این زمینه موجود است و آشنایی با ایده‌های نوآورانه آن‌ها.

مراجع

- https://fa.wikipedia.org/wiki/.DA/.AF/.DA/.AV/.D9/.A6/.DA/.B9/.DA/.AF/.DA/.AV/.D9/.AY_/.DA/.B9/.D9/.A4/.D9/.A5_/.D9/.AA_/.DA/.B5/.D9/.A6/.DA/.B9/.DA/.AA_/.DA/.AV/.DB/.AC/.DA/.B1/.DA/.AV/.D9/.A6 [1]
- http://ce.iust.ac.ir/index.php?sid=۱۴&slc_lang=fa [2]
- <https://www.crcv.ucf.edu/data/UCF۱۰۱.php> [3]