دوفصلنامه محاسبات و سامانه‌های توزیع شده
www.ijdcs.ir

سیستم پیشنهاد دهنده خرید تلفن همراه با استفاده از فیلترینگ دانش‌محور

رضا معروف[۱]، آرش خسروی*[۲]
[۱] دانشکده مهندسی برق و کامپیوتر، دانشگاه شهاب دانش ، قم
[۲] گروه مهندسی کامپیوتر، مرکز آموزش عالی محلات

چکیده

برای تسریع در انتخاب و تصمیم‌گیری صحیح در حوزه تلفن همراه، سیستم پیشنهاد دهنده تلفن همراه دانش محور همراه با پرسشنامه علایق فردی در این تحقیق ارائه شده است. در ابتدا داده‌های مشخصات فنی تلفن همراه گردآوری، پاکسازی، انتخاب و تبدیل شد. سپس با توجه به نیاز کاربران، برچسب‌هایی طبق چندین قانون استخراج و به داده‌ها اختصاص داده شد. پس از آن با استفاده از الگوریتم های طبقه بندی مانند درخت تصمیم، بیز ساده و مدل خطی تعمیم یافته یک مدل ترکیبی ارائه شد. این مدل ترکیبی همراه با الگوریتم های یادگیری عمیق و جنگل تصادفی به صورت جداگانه جهت طبقه‌بندی تلفن های همراه استفاده شد. در نهایت این مدل‌ها از لحاظ دقت ارزیابی شدند. کاربران می‌توانند با استفاده از مدل پیشنهادی و وارد کردن ترجیحات خود توسط پرسشنامه علایق فردی، به انتخاب تلفن همراه مناسب خود بپردازند.

کلمات کلیدی: سیستم پیشنهاد دهنده، دانش محور ، یادگیری عمیق

Mobile Recommender System using Knowledge based Filtering.

Reza Maroof[1], Arash Khosravi*[2]
[1]Faculty of Electrical and Computer Engineering, Shahab Danesh University, Qom
[2]Mahallat Institute of Higher Education, Mahallat, Iran

Abstract

In order to provide quick selection and decision making in the field of mobile phones marketing, a knowledge-based mobile recommender system along with a personal interest questionnaire is proposed in this research. Initially, the mobile technical data was collected, cleaned, selected, and converted, then assigned labels to the data according to several extracted rules according to users' needs. Then a hybrid model was proposed using classification algorithms such as decision tree, naive Bayes and generalized linear model. This hybrid model along with deep learning algorithms and random forest were used to classify cell phones separately. Finally, these models were evaluated in terms of accuracy. Users can choose their preferred mobile phone using the suggested model and enter their preferences using the Personal Interests Questionnaire.

تاریخچه مقاله:

تاریخ ارسال: ۹۸/۱۰/۰۱
تاریخ اصلاحات: ۹۸/۱۲/۰۱
تاریخ پذیرش: ۹۸/۱۲/۱۵
تاریخ انتشار: ۹۸/۱۲/۲۰

Keywords:

Recommender system
Knowledge based
Deep learning

روش ارجاع به مقاله:
سیستم پیشنهاد دهنده خرید تلفن همراه با استفاده از فیلترینگ دانش‌محور و پرسشنامه علایق فردی
ر.معروف، آ.خسروی، دو فصلنامه محاسبات و سامانه های توزیع شده، سال دوم، شماره دوم، شماره پیاپی ۴، سال ۱۳۹۸، ص ۱۱۰ تا ۱۳۰

Email: Khosravi.280@gmail.com

۱ – مقدمه

با افزایش روزانه داده‌ها, اطلاعات, کالاها در اینترنت مشکلاتی همچون افزونگی داده بوجود آمده است. مردم برای یافتن هدف مورد نظر خود باید زمان زیادی را صرف کنند. که حجم بالای اطلاعات آن‌ها را دچار مشکل و سردرگمی می‌کند و شاید هرگز به هدف مورد نظر و متناسب خود دست نیابند. یکی از راه‌حل های مقابله با افزونگی داده‌ها سیستم های پیشنهاد دهنده هستند. سیستم های پیشنهاد دهنده بوجود آمدند تا در کمترین زمان با شناخت از علایق افراد و داده‌ها بهترین و نزدیکترین گزینه به هدف مورد نظر کاربران را در میان انبوه داده‌ها و کالاها یافته و به آنها ارائه دهند. امروزه سیستم های پیشنهاد دهنده بسیار متداول شده‌اند. برای مثال در حوزه فیلم, موسیقی, کتاب‌ها، صنعت خوراکی، پزشکی، بهداشتی، مقاله‌ها, فروشگاه های آنلاین، گردشگری و شبکه های اجتماعی و... از سیستم های پیشنهاد دهنده استفاده می‌کنند. شرکت های مختلفی مانند آمازون[۱] برای فروش بیشتر و معرفی کالاهای خود از میان انبوه کالاها که روزانه در حال افزایش از هر نوع و مدل هستند، از سیستم پیشنهاد دهنده استفاده می‌کنند. این سیستم باعث می‌شود کاربر بدون اینکه درگیر جستجو در میان کالاهای مختلف شود، به سرعت به کالای هدف برسد. شرکت دیگری مانند نتفلیکس[۲] که در امر رسانه فعالیت می‌کند، برای اینکه کاربرانش زمان خود را برای جستجو در میان انبوه فیلم‌هایی که از محتوای آن اطلاع ندارند تلف نکنند، از سیستم پیشنهاد دهنده استفاده می‌کند. و به کاربرانش فیلم‌هایی متناسب علاقه آن‌ها معرفی می‌کند. بدیهی است اگر کاربر خود به جستجوی فیلم مورد علاقه خود بپردازد، ملزم به صرف وقت زیادی است که نتیجه نامعلومی نیز در پی دارد. با استفاده از سیستم های پیشنهاد دهنده نتایج فروش و بازدید آمازون و نتفلیکس بسیار افزایش داشته و امروز آمازون به یک عنصر جدایی ناپذیر در صنعت تجارت الکترونیک و خرده فروشی رسیده همینطور که نتفلیکس در صنعت رسانه بسیار مورد توجه است. سیستم های پیشنهاد دهنده تحولی در نتفلیکس ایجاد کردند که باعث شد نتفلیکس جایزه یک میلیون دلاری را برای تیمی در نظر بگیرد که بتواند ۱۰٪ سیستم پیشنهاد دهنده فعلی آنها را بهبود ببخشد. به همین دلایل ما نیز از سیستم های پیشنهاد دهنده یاری می‌جوییم تا در رسیدن به اهدافی که در نظر داریم به ما کمک کنند.

حال دلیل ما از انتخاب حوزه تلفن همراه این است که در ارتباط با خرید و انتخاب تلفن همراه سیستم های پیشنهاد دهنده بسیار اندک هستند. در صورتی که نیاز به تلفن همراه برای هر شخصی امری بدیهی شده است. طبق آمار بدست آماده از مرکز آمار ایران در سال ۱۳۹۶ حدود ۵۹ میلیون نفر از ۸۱ میلیون نفر جمعیت کل کشور ایران، کاربران تلفن همراه هستند. که حدود ۷۲٪ ضریب نفوذ را در میان مردم ایران نشان می‌دهد. به علت رشد فزاینده مدل های گوشی تلفن همراه در بازار و همچنین رشد استفاده اکثریت جامعه از تلفن همراه و عدم وجود اطلاعات کافی اکثریت افراد جامعه از سخت افزار و نرم افزار و مدل های مختلف تلفن های همراه, انتخاب تلفن همراه مناسب برای افراد دشوار

[۱] Amazon

[۲] Netflix

شده است. این انتخاب ملزم به صرف زمان بالا و تحقیقات بسیاری است تا هر فرد تلفن همراه مورد علاقه یا نیاز خود را پیدا کند. این امر اهمیت کار در این حوزه را نمایان می‌کند. صد البته این تمام دلیل نیست! تلفن های همراه ویژگی های بسیاری دارند که بسیار چالش برانگیز خواهد بود و اگر سیستم پیشنهاد دهنده قادر به تحلیل تلفن های همراه باشد آنگاه می تواند روی کالاها و داده‌های دیگر نیز عملکرد یکسان و حتی بهتری را ارائه کند.

هدف های ما در این مقاله به صورت زیر است:

- رفع مشکلات گذشته مانند محدودیت در ویژگی‌های تلفن همراه و ماتریس مقایسات زوجی که در عین سادگی قدرت انتخاب بیشتری به کاربران می‌دهد.
- تلفن های همراه طبقه بندی خواهند شد و ویژگی های آن‌ها استخراج و تبدیل خواهند شد و دانش مورد نیاز سیستم بدست خواهد آمد. این امر باعث می‌شود تلفن همراه هایی که برای اولین بار وارد سیستم می‌شوند بدون هیچ مشکلی در نتایج دخیل و مورد بررسی قرار بگیرند.
- نتایج فیترینگ دانش محور با پرسشنامه علاقه فردی در جهت افزایش رضایت کاربران از نتیجه نهایی با هم ادغام خواهد شد که باعث افزایش عملکرد می‌شود.
- نیاز های کاربران از تلفن همراه استخراج شده و به صورت یک بسته برای انتخاب به کاربران ارائه می‌شود. این کار باعث می‌شود کاربران مبتنی بر نیاز خود انتخاب را انجام دهند.
- برخی م شکلات سی ستم های گذ شته تحت عنوان شروع سرد وجود نخواهد داشت.
- سیستم پیشنهاد دهنده مبتنی بر اقلام ارائه خواهد شد که در سیستم‌های گذشته وجود ندارد. این امر باعث می‌شود محدودیت هایی که در سیستم مبتنی بر کاربران بود، برداشته شود.

۲ – ادبیات تحقیق

ریچی و همکاران [۱] بیان می‌کنند، سیستم پیشنهاد دهنده ابزار و رویکردی نرم افزاری است که با تحلیل کردن اقلام در راستای مورد استفاده قرار گرفتن توسط کاربران این اقلام را به کاربران پیشنهاد می‌کند. این پیشنهادات به کاربران در فرآیند تصمیم‌گیری های مختلف کمک می‌کنند. برای مثال سیستم پیشنهاد دهنده به آنها می‌گوید که چه جنسی را بخرند ، چه آهنگی را گوش دهند و یا چه خبرهایی را بخوانند. سیستم پیشنهاد دهنده به عنوان ارزشمند ترین ابزار جهت مقابله با سرریز داده ها شناخته شده و همچنین یکی از قوی‌ترین و محبوب‌ترین ابزار در زمینه تجارت الکترونیک است که به همین علت تکنیک های مختلفی در راه اندازی سیستم پیشنهاد دهنده در دهه های اخیر استفاده شده است. که بسیار هم در زمینه های تجاری موفق عمل کردند. توسعه یک سیستم پیشنهاد دهنده نیازمند تلاشی چند جانبه است. به این علت که حوزه های مختلفی را از جمله هوش مصنوعی، داده کاوی، آمار، بازاریابی، فناوری اطلاعات، مردم شناسی، دانش

تعامل انسان با کامپیوتر و پشتیبانی از تصمیم گیری را شامل می‌شود.

۱,۲ انواع سیستم های پیشنهاد دهنده

جاناچ و همکاران [۲] سیستم های پیشنهاد دهنده به چهار دسته اصلی تقسیم بندی می‌شوند.

۱- مبتنی بر فیلترینگ پالایش گروهی
۲- مبتنی بر فیلترینگ محتوا
۳- مبتنی بر فیلترینگ دانش محور
٤- مبتنی بر فیلترینگ ترکیبی

۱,۱,۲ پالایش گروهی

فلسفه فیلترینگ پالایش گروهی به این صورت می‌باشد که اگر کاربرانی با علایق و سلایق یکسان به یک مورد خاص علاقه داشته باشند، در این صورت در بقیه اقلام نیز ممکن است علایق مشترکی داشته باشند. همچنین کاربر دیگری با همان علایق و سلایق آنها هم به احتمال زیاد به آن مورد خاص و موارد دیگر مورد تایید و ترجیح آنها نیز علاقه خواهد داشت.

طبق تعریف ریچی و همکاران[۱] از فیلترینگ پالایش گروهی، منطق این است که اگر کاربر فعال(کاربری که پیشنهاد را دریافت می‌کند) در گذشته با برخی از کاربران در مواردی موافقت کند، دیگر پیشنهاد هایی که از طرف این کاربران مشابه می‌آید نیز باید مرتبط و مورد علاقه کاربر فعال باشد. در فیلترینگ پالایش گروهی مبنای کار ترجیحات کاربران است و این ترجیحات با امتیازدهی کاربران مشخص می‌شود. هر امتیاز اضافی که کاربران به اقلام مورد علاقه خود می‌دهند باعث شناخت بیشتر سیستم از کاربران شده و افزایش دقت پیشنهاد دهی را در پی دارد.

مهدی الاهی و همکاران [۳] نیز بیان می‌کنند که به طور کلی هرچه رتبه بندی و امتیازدهی از سوی کاربران بیشتر باشد پیشنهاد های سیستم بهتر و موثرتر خواهد بود.

۱,۱,۱,۲ مشکلات سیستم پالایش گروهی

شروع سرد: در دنیای واقعی تمامی اقلام و کاربران دارای اطلاعات کامل نیستند بسیاری از آنها اطلاعات و امتیاز های ناقصی دارند یا در مواردی اصلاً امتیاز دهی نشدند. آگاروال [۴] در سال ۲۰۱۶ بیان می‌کند که از مهم ترین چالش هایی که در طراحی یک سیستم پیشنهاد دهنده با فیلترینگ پالایش گروهی وجود دارد، می‌توان به نادر بودن ماتریس های رتبه بندی اولیه اشاره کرد. برای اینکه این چالش را بهتر درک کنیم دنیای فیلم را در نظر می‌گیریم. در این دنیا کاربران به فیلم هایی که مشاهده کردند امتیاز داده و فیلم هایی که مشاهده نشده اند امتیازی ندارند که شامل فیلم های بسیار زیادی نیز می‌شود. از این رو آن دسته از فیلم هایی که مشاهده نشده اند و امتیازی ندارند، در سیستم پیشنهاد دهنده جایی نخواهند داشت. اینگونه مشکلات را شروع سرد می‌گویند که در ابتدای کار اطلاعات یا ناقص است یا وجود ندارد.

پراکندگی داده: جاناچ و همکاران [۲] بیان می‌کنند که اگر بخواهیم از میان انبوه اطلاعات به دنبال هدف‌ها و علایق یکسان کاربران باشیم، این پراکندگی و افزونگی اطلاعات اغلب در دنیای واقعی استاندارد نبوده و نیازمند پاک‌سازی و نرمال‌سازی هستند. پس نیاز به قدرت

پردازش بسیار بالا و صرف زمان بسیاری دارند تا به یک داده صحیح و کارا تبدیل شوند.

۲،۱،۲ مبتنی بر محتوا

فیلترینگ مبتنی بر محتوا بر اساس انتخاب های گذشته کاربر عمل می‌کند. برای مثال یک وبسایت فروشگاه اینترنتی را در نظر بگیرید اگر فرض کنیم که در گذشته کاربر از این فروشگاه یک فیلم با ژانر خیالی خریداری کرده باشد آنگاه سیستم پیشنهاد دهنده فیلم هایی با داستان های خیالی اخیر را که کاربر خریداری نکرده به او پیشنهاد می‌کند. این سیستم با استفاده از اطلاعات از پیش تعیین شده برای هر یک از اقلام مانند کلید واژه هایی که برای اقلام در نظر گرفته می‌شود می‌تواند پیشنهادات خود را ارائه دهد. به معنای این که محتوا و یا خود اقلام قابل تجزیه و تحلیل هستند. مانند محتوای موسیقی، فیلم، عکس، نوشته ها و... . بنابراین می‌توان با استفاده از مقایسه شباهت میان اقلام به کاربرانی که اقلامی را مفید ارزیابی کردند اقلام مشابه آن را پیشنهاد کرد. برای مثال اگر کاربری یکی از اقلام را خریداری، لایک یا تماشا کرده، شنیده و یا امتیاز بالایی به آن داده است، این نوع از اقلام جزو پیشنهادات سیستم پیشنهاد دهنده مبتنی بر فیلترینگ محتوا به آن کاربر خاص خواهد بود[۵].

۳،۱،۲ دانش محور

سیستم پیشنهاد دهنده دانش محور با استفاده از اطلاعاتی که بطور مستقیم از کاربر دریافت می‌کند و اطلاعاتی که از اقلام دارد به ارائه پیشنهاد می‌پردازد. برای مثال اگر کاربر قصد خرید یک خانه را داشته باشد اطلاعاتی مانند متراژ، سن، قیمت و ... خانه را می‌تواند در ابتدای کار به سیستم ارائه کند تا با جستجو در میان اقلامی که دارای شرایط کاربر هستند بهترین آن‌ها را به کاربر پیشنهاد کند. این سیستم نیازی به امتیازدهی دیگر کاربران و بازخورد آن‌ها ندارد و بر خلاف سیستم های پیشنهاد دهنده قبلی از هیچگونه امتیاز و بازخورد دیگران استفاده نمی‌کند. این امر باعث شده تا در مواردی که اقلام در حال توسعه هستند بسیار کاربرد داشته باشد. برای مثال یک اتوموبیل با مدلی خاص ممکن است هر سال تکامل یابد که امتیازدهی آن را بسیار دشوار می‌کند. این اقلام خاص دارای ویژگی هایی مختص به خود هستند که کاربر فقط علاقمند به اقلامی با آن ویژگی است. برای مثال اتوموبیل ممکن است چندین مدل، نوع ساخت، رنگ، موتور و... داشته باشد و علاقه کاربر ترکیبی از این ویژگی‌ها باشد. در این صورت با استفاده از امتیازها و بازخوردها نمی‌توان به کاربر پیشنهاد خاص او را ارائه کرد. در این شرایط از سیستم پیشنهاد دهنده دانش محور استفاده می‌کنیم. حال که از امتیازات نمی‌توانیم استفاده کنیم، از نیازمندی های کاربران که در ابتدا در اختیار سیستم قرار می‌دهند و توضیحات و اطلاعات اقلام استفاده کرده و به مقایسه آن‌ها باهم میپردازیم تا بهترین پیشنهاد را به کاربر ارائه کنیم. سیستم پیشنهاده دهنده دانش محور به کاربران اجازه می‌دهد تا نیاز و علاقه خود را کاملاً مشخص کنند و سپس با پیمایش در بین اقلام مختلف خواسته کاربران را به آن‌ها پیشنهاد می‌کند[۴].

۴،۱،۲ سیستم ترکیبی

این نوع سیستم پیشنهاد دهنده از ترکیب سیستم های پیشنهاد دهنده با یکدیگر تشکیل می‌شود. به صورت کلی می‌توان گفت بهترین نوع سیستم پیشنهاد دهنده،

سیستم ترکیبی است. زیرا هر یک از سیستم‌ها در شرایط گوناگون مشکلات خاص خود را دارند. که با ترکیب کردن آنها می‌توان مشکلات آنها را حل کرد.

۳,۲ پیشینه تحقیق

در سال های گذشته سیستم های پیشنهاد دهنده مختلفی را پژوهشگران ارائه کردند. اما در حوزه تلفن همراه این سیستم‌ها بسیار اندک هستند و اکثریت مبتنی بر کاربران ارائه شدند. در این قسمت به بررسی چند مورد از سیستم های گذشته پرداخته و مشکلات آنها را بیان می‌کنیم.

سیستم پیشنهاد دهنده مبتنی بر فیلترینگ سلسله مراتبی در سال ۲۰۱۰ توسط چن و همکاران[۱۰] توسعه داده شده است. این سیستم ترجیحات و علایق کاربر را توسط یک ماتریس مقایسات زوجی از کاربر دریافت می‌کند و با استفاده از فیلترینگ سلسله مراتبی[۳] به وزن‌دهی به ویژگی های تعیین شده توسط کاربر می‌پردازد. در ابتدا ترجیحات کاربر را در بالا ترین سطر درخت سلسله مراتبی خود قرار می‌دهد و ویژگی های کم اهمیت‌تر را در پایین ترین سطح قرار می‌دهد و با تطبیق ویژگی‌های موبایل با درخت سلسله مراتبی ساخته شده بهترین پیشنهاد را به کاربر ارائه می‌کند. از ایرادات این سیستم می‌توان به اینکه کاربر موظف به انتخاب فقط یک ویژگی مهم برای خود هست اشاره کرد. به علاوه سیستم چن و همکاران فقط به بررسی تلفن همراه های کلاسیک پرداخته است. در تعریف ویژگی ها نیز به هیچ یک از ویژگی‌های تلفن های هوشمند نپرداخته و بسیار محدود عمل کرده است.

سیستم پیشنهاد دهنده زمینه آگاه تلفن همراه رضایی نور و لسانی[۱۱] با استفاده از ترکیب روش های تصمیم‌گیری جبرانی و غیر جبرانی در سال ۲۰۱۶ توسعه داده شده است. سیستم زمینه‌آگاه استفاده شده در این سیستم بدین معناست که زمان و مکان و عادات رفتاری کاربر را از روی تلفن همراه کاربر تحلیل و در جهت ارتقای میزان رضایت وی از آن استفاده می‌کند. این امر را بدین منظور مطرح کردیم تا به اولین مشکل این سیستم یعنی رعایت نکردن قوانین حریم خصوصی کاربر اشاره کنیم. با صرف نظر از مشکل حریم خصوصی دلیل پژوهشگران این سیستم برای استفاده از روش زمینه‌آگاه بدین منظور بوده است که هر کاربری علایق خاص خود را دارد. پس با استفاده از تحلیل اطلاعات کاربر، بدون پرسش و پاسخ از کاربر سعی در ارائه بهترین پیشنهاد به او را دارد. اما فرض کنیم کاربری علاقمند بازی باشد اما تلفن همراهی که در اختیار دارد و برای استخراج اطلاعات وی استفاده می‌شود قادر به اجرای بازی های مورد علاقه کاربر نباشد. آنگاه هیچگاه فرد نمی‌تواند به خواسته خود برسد و تلفن همراهی مناسب برای بازی را انتخاب کند. این سیستم اطلاعات زمینه‌ای کاربر را با استفاده از سنسور های تلفن همراه آن‌ها دریافت می‌کند و سپس با استفاده از سیستم چن و همکاران با بهره‌گیری از روش سلسله مراتبی آن به وزن‌دهی اطلاعات کاربر می‌پردازد. اگرچه این سیستم اطلاعات کاربر را تحلیل می‌کند اما در کنار آن از

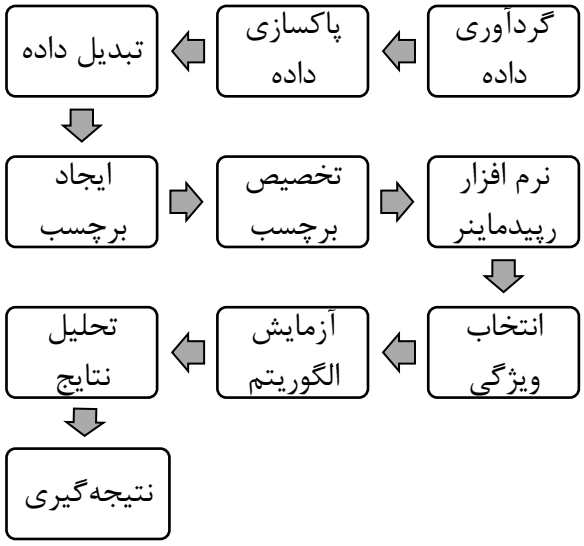
[۳] Hierarchy

ماتریس مقایسات زوجی چن و همکاران نیز با کمی تغییرات استفاده کرده است.

سیستم دیگری توسط بویوکوزکان و گلریوز[۱۲] در سال ۲۰۱۶ توسعه داده شده است که تمرکز آن بر روی یافتن میزان تفاوت هاست. در ابتدا به بررسی ویژگی های مختلف تلفن همراه اشاره کردند و سپس روش عملکرد تکنیک تاپسیس و روابط ریاضی مورد استفاده توضیح داده شده است. در ابتدا نیاز به ساخت یک ماتریس بیان شده است و باید با روش تاپسیس کار خود را به درستی انجام دهد. این تکنیک در صورتی که نزدیکی یا شباهتی به ترجیحات کاربر وجود داشته باشد از عدد یک که بالاترین شباهت هست تا عدد صفر که کاملا متناقض است، استفاده می‌کند. تمامی این عملکردها تحت نظریه فازی شهودی انجام شده است و در تمام قسمت ها نظریه فازی برقرار می باشد. در قسمت پیاده سازی نتایج را با استفاده از سه شخص متخصص در عرصه تلفن همراه در ترکیه آزمایش کردند. این آزمایش در گام اول با انتخاب سه مدل تلفن همراه که در بازار ترکیه هم قیمت هستند آغاز شد. سپس ویژگی های آن‌ها به صورت ماتریس استخراج شد. در گام دوم سه شخص مورد نظر علایق و ترجیحات خود را در اختیار سیستم قرار دادند تا سیستم ماتریس دیگری نیز برای این بخش بوجود آورد. این سیستم با استفاده از قوانین رتبه بندی لینگویستیک [۴] نظریه فازی شهودی را عملی و از آن استفاده می‌کند. چنانچه در این فرآیند تعداد تلفن های همراه زیاد باشد و تعداد کاربران نیز افزایش پیدا کند برای محاسبات در ساخت ماتریس ها و رتبه بندی و تشخیص شباهت ها ملزم به صرف زمان بسیار طولانی می‌باشد.

۳ – روش پژوهش

پس از معرفی سیستم‌های پیشنهاد دهنده موجود برای حل مشکلات گذشته، سیستم پیشنهاد دهنده دانش محور همراه با پرسشنامه علایق فردی را ارائه می‌کنیم. در شکل-۱ می‌توان خلاصه انجام پژوهش را مشاهده کرد.



(شکل-۱) خلاصه پژوهش

۱,۳ داده های تلفن همراه

قبل از هرچیز برای شروع نیاز به یک جدول داده تلفن همراه داریم. برای این منظور داده های یکی از معتبر ترین سایت های اطلاع رسانی و بررسی تلفن های همراه که در این حوزه ۲۰ سال سابقه داشته است، به نام Gsmarena.com را از انبار داده Kaggle گردآوری کردیم. این داده ها در بازه زمانی ۱۹۹۴ تا ۲۰۱۸ می‌باشند. جدول داده گردآوری شده اولیه دارای

[۴] Linguistic terms

۸۶۰۰ سطر و ۴۰ ستون بوده و دارای نواقص بسیاری است. به علت ناهمخوانی داده ها با الگوریتم‌ها، این ویژگی ها تبدیل به ستون های دیگری شده و به تعداد آنها در روند توسعه موتور دانش محور افزوده خواهد شد و به عدد ۷۲ خواهد رسید. اما از میان آنها ۳۵ ویژگی در مدل نهایی اهمیت خواهند داشت. به علاوه ستون های اضافه شده دیگری که برای کمک به بهبود پیشنهاددهی مندرج شدند. همچنین از میان سطرها نیز به صورت نظارت شده میزان ۲۰۱ سطر که نشان از مدل تلفن همراه های مختلف با قابلیت های متفاوت است، انتخاب کردیم. این ۲۰۱ تلفن همراه در بازار داخل کشور ایران در حال حاضر در دسترس هستند و در بازه زمانی ۲۰۱۶ تا ۲۰۱۸ روانه بازار ایران شده اند. در این ۲۰۱ سطر از شرکت های معروف تولید کننده تلفن همراه نظیر اَپل، سونی، سامسونگ، اِل‌جی، نوکیا، گوگل، وان‌پلاس[۵]، اِچ‌تی‌سی[۶]، هواوی[۷]، شیاومی[۸] و ایسوس استفاده شده است. عنوان هر ستون نیز به اختصار شامل برند، مدل، قیمت، صفحه نمایش، سیستم‌عامل، ریز پردازنده، دوربین و ... می‌باشد که در ادامه هریک را به صورت مشخص توضیح داده و چگونگی تبدیل آن‌ها به یک مقدار عددی قابل فهم برای الگوریتم را مشخص می‌کنیم.

۲,۳ انتخاب ویژگی‌های تلفن همراه

انتخاب ویژگی بر اساس سطح اهمیت ویژگی در یک تلفن همراه برای سیستم پیشنهاد دهنده بسیار مهم است. این اهمیت می‌تواند از نقطه نظر کاربران و یا از نقطه نظر عملکرد کلی سیستم از نظر دقت باشد. تلفن های همراه دارای ویژگی های بسیاری هستند که تنوع آنها یکی از چالش های این تحقیق بوده است. در این مقاله ویژگی های تلفن همراه را بطور کلی به سه دسته تقسیم کردیم. این سه دسته شامل ویژگی های بنیادی، سخت افزاری و نرم افزاری تلفن همراه می‌باشند.

۱. ویژگی های بنیادی

ویژگی های بنیادی شامل ساختار یک تلفن همراه و بهای مورد نیاز برای خرید آن است. از مهم‌ترین آن‌ها می‌توان به قیمت، برند، بدنه، زمان عرضه و صدا اشاره کرد.

۲. ویژگی های سخت افزاری

در یک تلفن همراه سخت افزار بسیار اهمیت دارد. در صورتی که سخت افزار یک تلفن همراه ضعیف باشد قادر به پاسخگویی به نیاز کاربران خود نخواهد بود و همچنین اگر بسیار قوی باشد ممکن است کاربران از لحاظ قیمت امکان خرید نداشته باشند. تلفن های همراه هوشمند فعلی قطعات سخت افزاری بسیاری دارند که هر یک برای انجام نوع خاصی از دستورات کاربران استفاده می‌شوند و البته در مجموع هر یک از این قطعات سخت افزاری باهم در ارتباط هستند. از جمله این قطعات می‌توان به واحد پردازش مرکزی[۹]، واحد پردازش گرافیکی[۱۰]،

[۵] OnePlus
[۶] HTC
[۷] Huawei
[۸] Xiaomi
[۹] Central Processing Unit
[۱۰] Graphical Processing Unit

حافظه دسترسی تصادفی[۱۱]، تراشه[۱۲]، صفحه نمایش، حافظه ذخیره سازی، باتری، دوربین و شبکه ها اشاره کرد.

۳. ویژگی های نرم افزاری

در قسمت های قبلی ویژگی های بنیادی و سخت افزاری تلفن همراه را بیان کردیم. اما هیچ یک از این ویژگی‌ها تا زمانی که نرم افزار نباشد عملکردی نخواهند داشت. وظیفه مدیریت تمام ویژگی‌ها بر عهده سیستم عامل است. در این میان به نسخه سیستم عامل نیز باید توجه شود. علاوه بر سیستم عامل، بعضی از تلفن های همراه قابلیت‌های نرم افزاری خاصی را ارائه می‌کنند. از این قابلیت‌ها می‌توان به دستیار صوتی و بازگشایی تلفن همراه از طریق تشخیص چهره اشاره کرد. برای مشخص شدن داده‌های هر ستون در جدول-۱ می‌توان نمونه داده‌های گردآوری شده را مشاهده کرد.

(جدول-۱): ستون ویژگی‌ها

نام ستون ویژگی‌ها	داده‌ها
Brand	Samsung
Model	Galaxy S8
Price	700 Euro
Announced	۲۰۱۷ March
Weight	155 g
SIM	Single Sim
Display Size	5.8 inches
Display Resolution	۱۴۴۰x 2960 pixels
Display Pixel Density	570 PPI
Operating System	Android Nougat
CPU	Octa-core (4x2.3 & 4x1.7) GHz
GPU	Mali-G71 MP20
Internal Memory	64 GB
External Memory	256 GB Supported
Ram	4 GB
Primary Camera	12 MP
Selfie Camera	8 MP
Camera Tech	Auto Focus-Phase Detection-OIS
Options	Iris Scanner-Finger Print-Gyro-NFC-…
Battery	3000 mAh

۳,۳ یکپارچه سازی و تصحیح نواقص

نقش اصلی در سیستم پیشنهاد دهنده دانش محور مربوط به داده ها است. در صورتی که داده اشتباه باشد و یا دارای نواقص مختلفی باشد بدیهی است که پیشنهاد اشتباهی را ارائه خواهد کرد. به همین منظور داده‌های گردآوری شده از هر نظر بررسی شدند تا به بالاترین میزان دقت در پیشنهاددهی دستیابیم. داده های گردآوری شده در قسمت های مختلفی دچار نقص های مختلفی بودند. بعضی از آن‌ها بدون مقدار[۱۳] بوده و بعضی دیگر مقادیر نامتعارف داشتند. که این نقص ها با مراجعه به وب سایت های شرکت های سازنده تلفن های همراه و منابع معتبر تلفن همراه به صورت دستی در قسمت دارای نقص درج و رفع شدند.

۴,۳ تبدیل داده ها

در جدول (۱-۳) نشان دادیم که داده‌های ما از نوع عددی نیستند. زمانی که داده‌ها عددی[۱۴] یا دودویی[۱۵] نباشند، بعضی از الگوریتم های طبقه بندی امکان پردازش این داده‌ها را نخواهند داشت. به همین منظور

[۱۱] Random Access Memory
[۱۲] Chipset
[۱۳] Missing Value
[۱۴] Numeric
[۱۵] Binary

شروع به تبدیل هر یک از ستون‌های جدول داده به داده‌های عددی کردیم.

سیستم عامل: سیستم عامل‌ها هریک توسط شرکت تولید کننده سیستم عامل نام گذاری شده و عددی به آنها اختصاص داده می‌شود. در جدول-۲ نحوه تبدیل نام هر نسخه از سیستم عامل نمایش داده شده‌است.

(جدول-۲): تبدیل نسخه سیستم عامل

نام نسخه سیستم عامل	عدد نسخه سیستم عامل	تبدیل شده
Jelly Bean	۴,۳-۴,۲-۴,۱	۴۰۱-۴۰۲-۴۰۳
Kit Kat	4.4	۴۰۴
Lollipop	+ 5.0	۵۰۰
Marshmallow	+ 6.0	۶۰۰
Nougat	+ 7.0	۷۰۰
Oreo	۸,۱-۸	۸۰۰-۸۰۱
Pie	۹	۹۰۰
Q	۱۰	۱۰۰۰

کارکرد: برای کمک به بهره وری بیشتر الگوریتم‌ها از آزمایش‌های سایت Benchmarks.ul.com استفاده کردیم و برای هر تلفن همراه میزان کارایی را در ستون جدیدی با نام کارکرد وارد کردیم. این امر به دلیل مکمل بودن چهار ویژگی ریزپردازنده، گرافیک، تراشه و حافظه تصادفی در کارکرد تلفن همراه انجام شد.

گرافیک: به علت تاثیر بسزای این ویژگی برروی ویژگی های دیگر، ویژگی گرافیک را به صورت اعداد طبیعی در جدول داده درج خواهیم کرد. این امر در مرحله اول با نرمال کردن داده‌ها آغاز می‌شود. برای مثال ممکن است تلفن های همراه از شرکت های مختلفی گرافیک تلفن همراه را تهیه کرده باشند، به این منظور قدرت هر یک نسبت به شرکتی که در ویژگی گرافیک داده فراوانی بیشتری دارد محاسبه خواهد شد. برای مثال در جدول داده فراوانی گرافیک‌های شرکت کوآلکام با مدل Adreno بسیار بیشتر از بقیه داده‌ها است. بنابراین دیگر داده‌ها را نسبت به عدد اختصاص داده شده توسط شرکت برای هر مدل، تبدیل می‌کنیم. برای مثال همانطور که در جدول-۳ مشاهده می‌کنید برای مدل Mali G71 نرمال‌سازی انجام شده‌است و عدد ۵۳۰ برای آن درنظر گرفته شده‌است.

(جدول-۳): تبدیل گرافیک

مدل گرافیک	شماره شرکت سازنده	عدد تبدیل شده نهایی
Adreno 530	۵۳۰	۵۳۰
Mali-G71 MP8	۷۱	۵۳۰
Adreno 510	۵۱۰	۵۱۰

دوربین: برای هر تکنولوژی به کار رفته در دوربین تلفن همراه یک امتیاز تلقی کرده و در قسمت مربوط به تکنولوژی دوربین مجموع آنها را قرار می‌دهیم. در جدول-۴ شرح تبدیلات دوربین تلفن همراه مشخص است.

(جدول-۴): تبدیل دوربین

ویژگی دوربین	داده	عدد اختصاصی
دقت سنسور	۱۲MP	۱۲
فاصله کانونی	f/1.7	1.7
تکنولوژی‌ها	Phase Detection-Autofocus-LED-OIS	4
تعداد دوربین	Dual Camera	2

۵،۳ برچسب‌ها

الگوریتم های طبقه بندی استفاده شده در این پایان نامه از نوع یادگیری با نظارت هستند. به این معنا که نیاز به یک نظارت برای تصمیم‌گیری در مسائل برای آن‌ها وجود دارد. این نظارت می‌تواند با برچسب زدن بر روی یک سطر از داده‌ها صورت پذیرد. متخصص حوزه مسئله، مسئله را بررسی کرده و نظر خود را به صورت برچسب بر روی سطر مورد نظر اعمال می‌کند.

نکته مهم در این قسمت در ابتدا توجه به نیاز کاربران است و سپس طبقه‌بندی صحیح تلفن همراه حائز اهمیت خواهد بود. این برچسب‌ها باید نیاز و علایق کاربران را به خوبی پوشش دهند. عنوان برچسب‌ها شامل برچسب حرفه‌ای، مخصوص بازی، رسانه[۱۶]، اجتماعی[۱۷]، انرژی، به صرفه اقتصادی[۱۸]، مخصوص عکاسی، سلفی، آپشنال[۱۹] ، جدید، دو سیم کارت، کلاسیک و سطح پایین می‌باشد. لازم به ذکر است از برچسب‌های آپشنال، جدید و دو سیم کارت برای پیشبینی برچسب‌های دیگر استفاده شده است.

در حوزه تلفن های همراه می‌توان قوانین مختلفی را به نسبت سخت‌افزار و نرم‌افزار استفاده شده برای هر یک از تلفن های همراه استخراج کرد. با کمک گرفتن از این قوانین و نظرات افراد متخصص فعال در حوزه تلفن همراه، فرآیند برچسب زدن با نظارت را انجام می‌دهیم.

در پایان این فرآیند تمام تلفن های همراه جدول داده دارای یک برچسب خواهند بود. که این برچسب نمایان‌گر طبقه آن تلفن همراه است.

در (جدول-۵) قوانین و معیارهایی که در هر برچسب دخیل می‌باشند را می‌توان مشاهده کرد.

(جدول-۵) قوانین برچسب‌ها

نام برچسب	قوانین
Dual Sim	IF Sim >= 2 Then True.
Economic	IF Price < 200 && Performance>4000 Then True.
Energy	IF Battery >4000&&Display PPI <450 Then True.
Gaming	IF Performance >25000 Then True.
Multimedia	IF Display Size>=5.5 && Display Resolution>=1080&& GPU>=500 && Battery>3000 && External Memory>32 Then True.
New	IF Announced Time > 2017.03 Then True.
Photography	IF (Camera +Camera Sensor *1000)>25000 Then True.

[۱۶] Multimedia
[۱۷] Social
[۱۸] Economic
[۱۹] Optional

Professional	IF At Least Have Contain 4 Other Labels With High Level.
Social	IF Weight <170 &&Display Size Between 6-4&&Display Resolution>=720&&Internal Memory>=16 Then True.
Selfie	IF Selfie Camera>5&&GPU>400 Then True.
Optional	IF Sensor number >=3 && Finger number==1 Or NFC number==1 && Type C number==1 Then True.

۶,۳ پرسشنامه علایق فردی

برای اینکه بتوانیم بهترین پیشنهاد را به کاربران ارائه کنیم، باید از خواسته و علایق آن‌ها با خبر باشیم. در این پرسشنامه به صورت کاملا صریح از کاربر خواسته می‌شود تا نیاز و علاقه خود را اولویت بندی کند. البته کاربر در هر لحظه از فرآیند ارائه پیشنهاد می‌تواند، اولویت های خود را تغییر دهد. از آنجا که یک تلفن همراه می‌تواند چندین ویژگی را داشته باشد، چهار سطح اولویت برای کاربران در نظر گرفته می‌شود. شکل-۲ تعامل کاربر با سیستم پیشنهاد دهنده را نمایش می‌دهد.

(شکل-۲) تعامل کاربر با سیستم پیشنهاد دهنده

همانطور که در شکل نمایانگر بود کاربر ترجیحات خود را از کادر مورد نظر انتخاب می‌کند. اولویت‌های کاربر از میان ویژگی‌هایی است که در جدول داده برای تلفن های همراه تعریف شده است. برای مثال کاربر با اولویت قرار دادن صفحه نمایش بزرگ و باتری در لیست اولویت‌های اول و دوم خود اشاره به برچسب رسانه می‌کند. سیستم پیشنهاد دهنده این نیازمندی را دریافت کرده و شروع بر پیمایش در میان تلفن های همراه با این مشخصات می‌کند. از آنجایی که قیمت مد نظر و اولویت سوم کاربر مقرون به صرفه است، سیستم سعی در جستجوی تلفن های همراه با قیمت مناسب و دارای شرایط درج شده می‌کند. حال میتوان در شکل-۳ پیشنهادات ارائه شده به کاربر را مشاهده کرد.

(شکل ۳-۳) ارائه پیشنهادات توسط سیستم

اگر به نتایج ارائه شده توسط سیستم پیشنهاد دهنده و به اولویت های کاربر دقت کنیم، اولویت‌ها و نتایج تقریباً به برچسب رسانه اشاره می‌کند. و سیستم پیشنهاد دهنده نتایج را از طبقه رسانه انتخاب کرده است.

۷,۳ معماری سیستم

سیستم پیشنهاد دهنده دانش محور در مرحله اول درخواست ثبت اطلاعات توسط کاربر را ارسال می‌کند. کاربر به قسمت پرسشنامه راهنمایی می‌شود تا اولویت‌های خود را ثبت کند. حال در مرحله دوم نوبت به سیستم می‌رسد تا نیازهای کاربر را تشخیص دهد.

سیستم پیشنهاد دهنده دانش محور ارائه شده در این پایان نامه مبتنی بر اقلام است. یعنی تمرکز سیستم بر روی استخراج اطلاعات از اقلام است و اطلاعات کاربر را مستقیم از خود او توسط پرسشنامه دریافت می‌کند. اطلاعات اقلام توسط الگوریتم های طبقه‌بندی بررسی و به هر کدام برچسب اختصاص داده خواهد شد و دانش مورد نیاز تشخیص اقلام استخراج می‌شود. سپس با اتکا بر دانش بدست آورده توسط الگوریتم‌های طبقه‌بندی سیستم شروع به جستجو بر مبنای اولویت های کاربر می‌کند. این اطلاعات از قبل در پایگاه داده ذخیره شده و با دستور کاربر فرآیند جستجو آغاز می‌شود. در مرحله سوم تمام اطلاعات تلفن های همراه و کاربران به موتور پردازش دانش محور انتقال داده می‌شود تا تصمیم‌گیری نهایی انجام شود. پس از اتمام پردازش در مرحله سوم، پیشنهادات به کاربر اعلام می‌شوند. در صورتی که کاربر پیشنهاد مورد نظر را انتخاب کند فرآیند به پایان رسیده و در صورتی که پیشنهادات مورد پسند کاربر نباشد، می‌تواند اولویت های خود را تغییر دهد و پیشنهادات جدیدی دریافت کند. در این میان موارد پیشنهادی در پروفایل کاربر ذخیره شده و پیشنهادات قبلی که مورد پسند کاربر نبوده را ارائه نخواهد داد. مراحل توضیح داده شده در ارتباط با معماری سیستم در شکل-۴ مشاهده کرد.

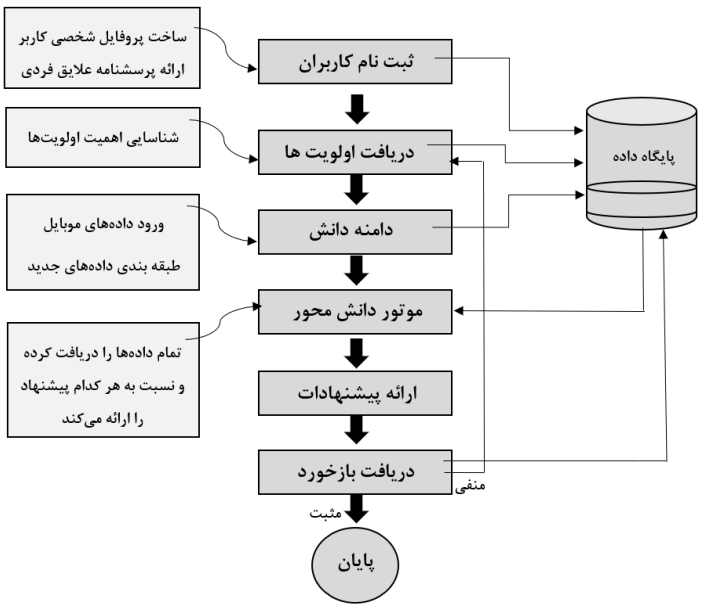
شکل-۴ – تجزیه و تحلیل داده ها

در این بخش به تجزیه و تحلیل نتایج بدست آمده توسط الگوریتم‌های مختلف می‌پردازیم.

۱,۴ مدل ترکیبی

انجام عملیات طبقه‌بندی توسط مدل ترکیبی بر روی برچسب Last Label انجام می‌شود. این برچسب خود دارای برچسب‌های مختلفی است که هر کدام

نشان‌دهنده نوع و طبقه تلفن همراه است. جزییات دقت طبقه‌بندی مدل ترکیبی را برای هریک از برچسب‌ها می‌توان در جدول-۶ مشاهده کرد.



(شکل-۴)

همانطور که در جدول-۶ مشاهده می‌کنید، ما برای سنجش مدل‌ها از سه نوع معیار اندازه‌گیری استفاده می‌کنیم. معیار اول که برای ما بیشتر مورد اهمیت است، معیار دقت است. معیار دوم صحت[۲۰] است و معیار سوم بازخوانی[۲۱] است. دقت مدل ترکیبی برابر با ۸۶,۵٪ است. که از نظر دقت نسبت به مدل‌های دیگری که با الگوریتم‌های بیز ساده، درخت تصمیم و مدل خطی تعمیم یافته به صورت جداگانه انجام شده، بسیار بالاتر است. اما این یک دقت کلی است. اگر به نتایج جدول در انتها برای هر یک از ویژگی‌ها دقت کنیم، خواهیم دید که درصد بازخوانی در بعضی موارد ۶۰٪ رسیده است و در بعضی دیگر به بالای ۹۰٪ می‌رسد.

۲,۴ جنگل تصادفی

(جدول-۶) جزییات دقت طبقه‌بندی مدل ترکیبی

Accuracy 86.57 %	true Pro	true Multi	true General	true Energy	true Eco	true Gaming	true Photo	true Low	true Classic	class precision
pred. Pro	23	1	0	0	0	1	1	0	0	88.46%
pred. Multi	1	16	0	0	0	0	2	0	0	84.21%
pred. General	0	1	62	0	7	0	1	1	0	86.11%
pred. Energy	0	0	0	27	0	0	0	0	0	100.00%
pred. Eco	0	0	2	0	12	0	0	0	0	85.71%
pred. Gaming	1	0	0	0	0	8	2	0	0	72.73%
pred. Photo	1	1	1	0	0	2	9	0	0	64.29%
pred. Low	0	0	0	0	1	0	0	14	0	93.33%
pred. Classic	0	0	0	0	0	0	0	0	3	100.00%
class recall	88.46%	84.21%	95.38%	100.00%	60.00%	72.73%	60.00%	93.33%	100.00%	

[۲۰] Precision

[۲۱] Recall

پس از اجرا دقت این مدل برابر با ۸۳,۵٪ بوده، که از تک الگوریتم‌های قبلی بالاتر و از ترکیب آن‌ها پایین‌تر است. در جدول-۷ دقت طبقه‌بندی برای هر ویژگی را می‌توان مشاهده کرد.

واقعیت غلط نیست و باعث می‌شود که میزان بازخوانی در بعضی موارد کم شود. نکته دیگر چند برچسبی بودن است که باعث کم شدن میزان بازخوانی می‌شود. برای مثال در پیشبینی برچسب Eco میزان بازخوانی برابر با ۴۰٪ است که عدد خوبی نیست.

(جدول-۷) جزییات دقت طبقه‌بندی جنگل تصادفی

Accuracy 83.5 %	true Pro	true Multi	true Social	true Energy	true Eco	true Gaming	true Photo	true Low	true Classic	class precision
pred. Pro	23	1	0	0	0	2	0	0	0	88.46%
pred. Multi	0	15	0	0	0	0	1	0	0	93.75%
pred. Social	0	2	63	0	12	0	2	1	0	78.75%
pred. Energy	2	0	0	27	0	0	0	0	0	93.10%
pred. Eco	0	0	0	0	8	0	0	0	0	100.00%
pred. Gaming	1	0	0	0	0	6	3	0	0	60.00%
pred. Photo	0	1	1	0	0	3	9	0	0	64.29%
pred. Low	0	0	1	0	0	0	0	14	0	93.33%
pred. Classic	0	0	0	0	0	0	0	0	3	100.00%
class recall	88.46%	78.95%	96.92%	100.00%	40.00%	54.55%	60.00%	93.33%	100.00%	

الگوریتم جنگل تصادفی دقت ۸۳,۵٪ را ارائه می‌دهد. این الگوریتم دارای میزان صحت خوبی است اما در بعضی موارد میزان بازخوانی این الگوریتم بسیار پایین بوده است.

البته باید به این نکته توجه داشت که در حوزه پیشنهاددهی تلفن همراه میزان بازخوانی پایین مشکل بزرگی حساب نمی‌شود. زیرا بعضی از تلفن های همراه چندین مشخصه را باهم دارند و این پیشبینی در

۳,۴ یادگیری عمیق

مدل مبتنی بر الگوریتم یادگیری عمیق دقتی برابر با ۸۶٪ را ارائه می‌دهد. که از تمام الگوریتم‌های که به صورت جداگانه مدل شده باشند، دقت بالاتری را ارائه می‌کند. اما در مقایسه با مدل ترکیبی دقت کمتری دارد. دقت مدل مبتنی بر الگوریتم یادگیری عمیق را در جدول-۸ می‌توان مشاهده کرد.

جدول-۸ جزییات دقت طبقه‌بندی یادگیری عمیق

Accuracy 86.02 %	true Pro	true Multi	true General	true Energy	true Eco	true Gaming	true Photo	true Low	true Classic	class precision
pred. Pro	22	0	0	0	0	2	0	0	0	91.67%
pred. Multi	1	14	0	0	0	0	1	0	0	87.50%
pred. General	0	3	62	1	5	0	2	1	0	83.78%
pred. Energy	0	1	0	26	0	0	0	0	0	96.30%
pred. Eco	0	0	2	0	15	0	0	0	0	88.24%
pred. Gaming	3	1	0	0	0	6	1	0	0	54.55%
pred. Photo	0	0	1	0	0	3	11	0	0	73.33%
pred. Low	0	0	0	0	0	0	0	14	0	100.00%
pred. Classic	0	0	0	0	0	0	0	0	3	100.00%
class recall	84.62%	73.68%	95.38%	96.30%	75.00%	54.55%	73.33%	93.33%	100.00%	

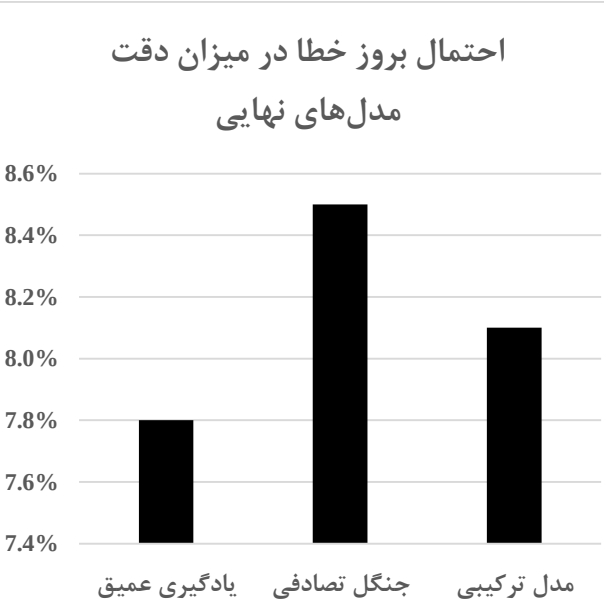
همانگونه که در قسمت‌های قبلی مشاهده شد، دقت یادگیری عمیق کمتر از مدل ترکیبی بوده است. اما نکته حائز اهمیت در این مدل وجود تعادل بهتر در نتایج هر یک از ویژگی ها است. دقت، صحت، بازخوانی هر سه در این مدل تعادل خوبی دارند و هیچکدام به حدی کم نیست که ایجاد مشکل کند. لازم به ذکر است که تعداد داده‌ها برای الگوریتم یادگیری عمیق کم بوده است. اما با این اوصاف عملکرد خوبی را داشته است. تمام ویژگی ها دقت خوبی را ارائه می‌کنند و با بیشتر شدن تعداد داده‌ها این روند ادامه خواهد داشت و حتی بهتر نیز خواهد شد.

۴,۴ احتمال بروز خطا

برای اینکه بتوانیم بهترین الگوریتم طبقه‌بندی را برای تلفن همراه مشخص کنیم، معیار دیگری به نام احتمال بروز خطا را نیز باید برای هر سه الگوریتمی که بیشترین دقت را ارئه کردند، بررسی کنیم.

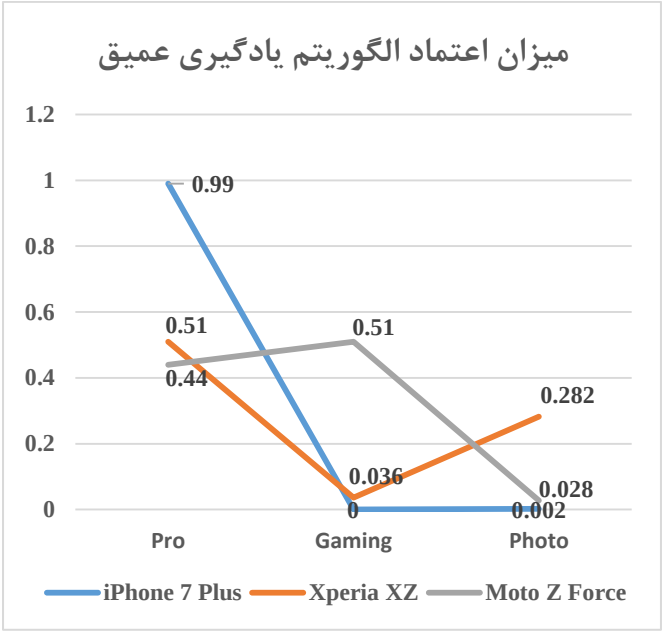
این احتمال بیان می‌کند که محتمل است دقت به میزان چندین درصد بالاتر یا پایین‌تر قرار بگیرد. احتمال بروز خطا در دقت هرچه کمتر باشد، بهتر است و نتیجه دقیق‌تری را ارائه می‌کند.

احتمال بروز خطا در هر یک از مدل های نهایی را می‌توان در شکل-۴ مشاهده کرد.



(شکل-۴) احتمال بروز خطا در مدل‌های نهایی

جنگل تصادفی با ۸,۵٪ بیشترین احتمال بروز خطا را در دقت دارد. پس از آن مدل ترکیبی با ۸,۱٪ و کمترین میزان احتمال بروز خطا را یادگیری عمیق با ۷,۸٪ دارد. برای تحلیل بهتر مفهوم خطا در طبقه‌بندی تلفن همراه و دلیل انتخاب معیار دقت برای اندازه‌گیری، می‌توان در شکل-۵ نتایج پیشبینی شده برای سه تلفن همراهی که برچسب Pro به آنها اختصاص داده شده است را مشاهده کرد. تصمیم‌گیری الگوریتم بر اساس معیار میزان اعتماد است. میزان اعتماد[۲۲] معیار تشخیص الگوریتم در تعلق یک برچسب به تلفن همراه است. این معیار از عدد صفر که کمترین میزان و عدد یک که بالاترین میزان اعتماد است، محاسبه می‌شود. مثالی از خطای طبقه‌بندی الگوریتم یادگیری عمیق با استفاده از قابلیت اعتماد را می‌توان در شکل-۵ مشاهده کرد.



(شکل-۵) میزان اعتماد یادگیری عمیق

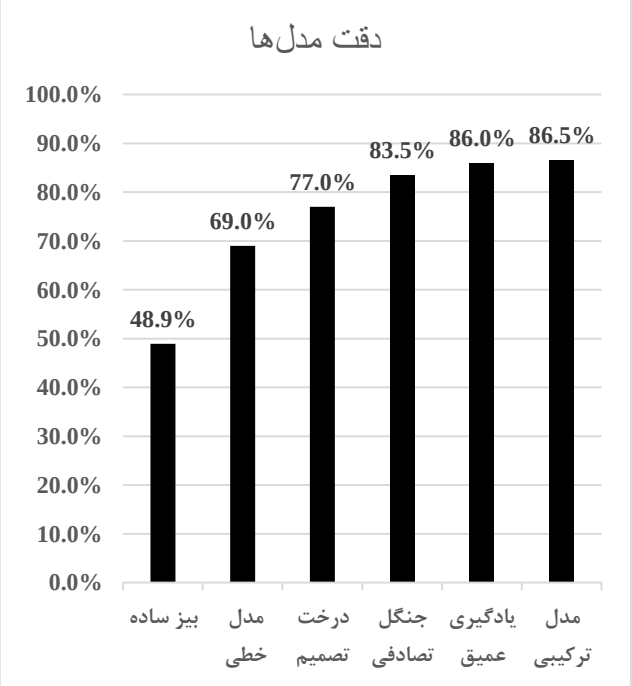
همانطور که در شکل-۵ قابل مشاهده است، برچسب هر سه تلفن همراه Pro بوده که برای تلفن همراه iPhone 7 Plus با بالاترین میزان اعتماد درست پیشبینی شده است. اما برای تلفن همراه Xperia XZ این میزان اعتماد کمتر است اما همچنان درست پیشبینی شده است. و برای تلفن همراه Moto Z Force با اختلاف کمی اشتباه پیشبینی شده‌است. اما این اشتباه در حقیقت یک اشتباه محاسباتی است. در حقیقت یک تلفن همراه می‌تواند به برچسبی که نزدیک‌ترین سطح را به سطح برچسب ناظر دارد تعلق داشته باشد. میزان اعتماد تعلق بیشتر یک تلفن همراه را به برچسب پیش‌بینی شده نمایان می‌کند که حتی می‌توان از میزان اعتماد برای ارائه پیشنهادهای چند

[۲۲] Confidence

برچسبی استفاده کرد. البته لازم به ذکر است برای تلفن های همراه با برچسب Pro صدق می‌کند.

۵- نتیجه‌گیری و پژوهش‌های آتی

نتیجه‌گیری نهایی بر اساس دقت مدل‌های ارائه شده در این مقاله است. همانطور که اشاره شد دقت مدل ترکیبی از دیگر مدل‌ها بالاتر بوده است. اگر از جنگل تصادفی استفاده کنیم ممکن است با افزایش تعداد داده‌ها نیاز به زمان بیشتری برای طبقه‌بندی داشته باشیم. یادگیری عمیق نیز دقت خوبی را ارائه کرده که ممکن است در شرایط دیگری نتیجه بهتری را ارائه دهد. اما در حال حاضر مدل ترکیبی با اختلاف کمی دقت بالاتری را ارائه کرده است. در شکل-۶ می‌توان دقت مدل‌ها را مشاهده کرد.



(شکل -۶) دقت تمامی مدل‌ها

با تمام این اوصاف نتیجه می‌گیریم اگر جدول داده سالمی از تلفن های همراه را در اختیار داشته باشیم که با استفاده از مدل ترکیبی طبقه‌بندی شده باشد، آنگاه در صورتی که کاربران خواسته های خود را به درستی وارد پرسشنامه کرده باشند، می‌توان به رضایت ۸۶ درصدی کاربران از پیشنهاد ارائه شده به آن‌ها دست یافت. البته به احتمال بروز خطا در دقت نیز باید توجه داشت.

طبق نتایج بدست آمده از آزمایش مدل‌ها و تحلیل داده‌ها، برای جدول داده استفاده شده در این مقاله مدل ترکیبی انتخاب بهتری است. چرا که این الگوریتم برای برچسب‌ها دقت پیشبینی بالاتری را ارائه می‌کند. اگرچه اختلاف کمی با دیگر الگوریتم‌ها دارد. اما اگر بخواهیم جدول داده خود را توسعه دهیم و تعداد بسیار بیشتری از تلفن های همراه را طبقه‌بندی کنیم، آنگاه الگوریتم یادگیری عمیق دقت بالاتر و نتایج بهتری را ارائه می‌کند. این الگوریتم در میزان صحت و بازخوانی نیز تعادل بسیار خوبی را برقرار می‌کند. همچنین درصد خطای این الگوریتم از دیگر الگوریتم‌ها بسیار کمتر بوده که نکته قابل توجهی است. بنابراین کاربران این سیستم پیشنهاد دهنده می‌توانند بدون نیاز به داشتن اطلاعاتی از تلفن های همراه فقط با تکیه بر دانش بوجود آمده از الگوریتم یادگیری عمیق بر روی داده‌های تلفن همراه، براحتی با کمترین خطا به هدف مورد نظر خود در حوزه تلفن همراه برسند. این سیستم برای سایت های تجاری حوزه تلفن همراه نیز بسیار مفید خواهد بود. زیرا مشتریان آنها می‌توانند با صرف کمترین زمان تلفن همراه مورد نیاز خود را انتخاب و خرید را انجام دهد. بدلیل اینکه سیستم پیشنهاد دهنده دانش محور ارائه شده متمرکز بر اقلام بوده است، بنابراین برای طبقه‌بندی تلفن های همراه در یک سایت تجاری نیز

می‌توان از سیستم پیشنهاد دهنده ارائه شده در این مقاله کمک گرفت.

در این مقاله بیشتر تمرکز بر روی تلفن همراه بوده تا بتوان اقلام جدید را به راحتی طبقه‌بندی و به کاربران پیشنهاد کرد. اما از لحاظ شناخت کاربران تلاشی صورت نگرفته و فقط به صورت صریح از کاربر نیاز و علایق را توسط پرسشنامه دریافت کرده ایم. همچنین می‌توان با استفاده از ترکیب سیستم پیشنهاد دهنده دانش محور و سیستم پیشنهاد دهنده پالایش گروهی به نتایج بهتری نیز دست یافت. که این عمل نیازمند جمع آوری نظرات کاربران مختلف در ارتباط با هر تلفن همراه است. در بخش داده‌ها به علت کمبود بعضی از داده‌ها در عرصه تلفن همراه عواملی مانند مقاومت در برابر ضربه و کیفیت ساخت لحاظ نشده است. که در صورتی که این اطلاعات قابل دستیابی باشد، می‌توان به کاربران علاقمند این حوزه نیز پیشنهادهای مناسبی را ارائه کرد.

۶ – مراجع

[1] Ricci, Francesco, Lior Rokach, and Bracha Shapira. "Introduction to recommender systems handbook." In Recommender systems handbook, pp. 1-35. Springer, Boston, MA, 2011.

[2] Jannach, Dietmar, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich. "An introduction to recommender systems." New York: Cambridge (2011).

[3] Jannach, Dietmar, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich. "An introduction to recommender systems." New York: Cambridge (2011).

[4] Elahi, Mehdi, Francesco Ricci, and Neil Rubens. "A survey of active learning in collaborative filtering recommender systems." Computer Science Review 20 (2016): 29-50.

[5] C. Aggarwal, Recommender systems: The Textbook, Cham: Springer International Publishing, 2016.

[6] J. Bobadilla, F. Ortega, A. Hernando, A. Gutiérrez, Recommender systems survey, Knowledge-Based Systems: 109-132, 2013.

[7] J. Han, M. Kamber, J. Pei, Data Mining Concepts and Techniques Third Edition, The Morgan Kaufmann Series in Data Management Systems, 2011.

[8] A. J. Dobson, A. G. Barnett, An introduction to generalized linear models. Chapman and Hall/CRC, 2008.

[9] L. Deng, D. Yu, Deep Learning Methods and Applications, 2014.

[10] Y. Lecun, Y. Bengio, G. Hinton, Deep Learning, 2015.

[11] Chen, Deng-Neng, Paul Jen-Hwa Hu, Ya-Ru Kuo, and Ting-Peng Liang. "A Web-based personalized recommendation system for mobile phone selection: Design, implementation, and evaluation." Expert Systems with Applications 37, no. 12 (2010): 8201-8210.

[12] Rezaeenour, Jalal, and Fatemeh Sadat Lesani. "A Context Aware Recommender System for Mobile Phone Selection Using Combination of Elimination Method and Analytic Hierarchy Processing." Iranian Journal of Information processing and Management 32, no. 4 (2017): 1203-1228.

[13] Büyüközkan, Gülçin, and Sezin Güleryüz. "Multi criteria group decision making approach for smart phone selection using intuitionistic fuzzy TOPSIS." International Journal of Computational Intelligence Systems 9, no. 4 (2016): 709-725.

[14] Smith, Brent, and Greg Linden. "Two decades of recommender systems at amazon. com." Ieee internet computing 21, no. 3 (2017): 12-18.

[15] Gomez-Uribe, Carlos A., and Neil Hunt. "The netflix recommender system: Algorithms, business value, and innovation." ACM Transactions on Management Information Systems (TMIS) 6, no. 4 (2015): 1-19.

[16] Covington, Paul, Jay Adams, and Emre Sargin. "Deep neural networks for youtube recommendations." In Proceedings of the 10th ACM conference on recommender systems, pp. 191-198. 2016.

رضا معروف مدرک کارشناسی خود را در رشته مهندسی نرم افزار در سال ۱۳۹۵ از دانشگاه آزاد ساوه اخذ کرده است. ایشان در حال حاضر دانشجوی کارشناسی ارشد در رشته مهندسی نرم افزار در دانشگاه شهاب دانش قم هستند.

آرش خسروی مدرک کارشناسی خود را در رشته مهندسی نرم افزار در سال ۱۳۸۲ از دانشگاه صنعتی اصفهان، مدرک کارشناسی ارشد خود را در رشته مهندسی فناوری اطلاعات در سال ۱۳۹۲ و مدرک دکتری خود را در رشته مهندسی فناوری اطلاعات، گرایش سیستم های اطلاعاتی در سال ۹۶ از دانشگاه صنعتی مالزی اخذ کرده است. ایشان در حال حاضر به عنوان هیات علمی مرکز آموزش عالی محلات مشغول به کار هستند. زمینه های پژوهشی مورد علاقه ایشان عبارتند از: هوش تجاری، سیستم های پیشنهاد دهنده، مدیریت دانش مشتری، داده کاوی، متن‌کاوی، فناوری اطلاعات در پزشکی.