بسم الله الرحمن الرحيم

**دانشکده مهندسی کامپیوتر**
**رساله دکتری مهندسی هوش مصنوعی**

# ارائه سیستم توصیه‌گر خبر مبتنی بر تمایلات کاربر

**نگارنده: باقر رحیم‌پورکامی**

**استاد راهنما**
**دکتر حمید حسن‌پور**

**استاد مشاور**
**دکتر هدی مشایخی**

شهریور ۱۳۹۷

شماره: ۵۸ ف.ت
تاریخ: ۹۷/۷/۸
ویرایش:

**باسمه تعالی**

دانشگاه صنعتی شاهرود
مدیریت تحصیلات تکمیلی

**فرم شماره ۱۲: صورت جلسه نهایی دفاع از رساله دکتری (Ph.D)**

(ویژه دانشجویان ورودی های ۹۴ و ما قبل)

بدینوسیله گواهی می شود آقای **باقر رحیم‌پور کامی** دانشجوی دکتری رشته **مهندسی کامپیوتر-هوش مصنوعی** به شماره دانشجویی **۹۲۳۳۲۰۵** ورودی مهرماه سال ۱۳۹۲ در تاریخ ۱۳۹۷/۰۶/۱۹ از رساله **نظری**☐ / **عملی**■ خود با عنوان : **ارائه سیستم توصیه‌گر خبر مبتنی بر تمایلات کاربر** دفاع و با اخذ نمره ۱۷/۳۶ به درجه : بسیار خوب نائل گردید.

| | |
|---|---|
| الف) درجه عالی: نمره ۲۰-۱۹ ☐ | ب) درجه بسیار خوب: نمره ۱۸/۹۹ – ۱۷ ☒ |
| ج) درجه خوب: نمره۱۶/۹۹– ۱۵ ☐ | د) غیر قابل قبول و نیاز به دفاع مجدد دارد ☐ |
| ه) رساله نیاز به اصلاحات دارد ☐ | |

| ردیف | هیئت داوران | نام و نام خانوادگی | مرتبه علمی | امضاء |
|---|---|---|---|---|
| | استاد راهنما | آقای دکتر حمید حسن‌پور | استاد | |
| | استاد مشاور | خانم دکتر هدی مشایخی | استادیار | |
| | داور خارجی | آقای دکتر بهروز مینایی | دانشیار | |
| | داور داخلی | آقای دکتر مرتضی زاهدی | استادیار | |
| | داور داخلی | آقای دکتر علیرضا احمدی فرد | دانشیار | |
| | نماینده تحصیلات تکمیلی | آقای دکتر منصور فاتح | استادیار | |

---

**مدیر محترم تحصیلات تکمیلی دانشگاه:**

ضمن تأیید مراتب فوق مقرر فرمائید اقدامات لازم در خصوص انجام مراحل دانش آموختگی آقای باقر رحیم‌پور کامی بعمل آید.

**نام و نام خانوادگی رئیس دانشکده :**

**تاریخ و امضاء و مهر دانشکده:**

وزارت علوم، تحقیقات و فناوری
دانشکده فناوری اطلاعات و کامپیوتر

شماره: ۵۸۴/ف ک
تاریخ: ۹۷/۷/۸
ویرایش:

**باسمه تعالی**

**مدیریت تحصیلات تکمیلی**

**پیوست شماره ۲**

**دانشکده : مهندسی کامپیوتر**

**گروه : هوش مصنوعی**

رساله دکتری آقای باقر رحیم‌پورکامی

**تحت عنوان: ارائه سیستم توصیه‌گر خبر مبتنی بر تمایلات کاربر**

در تاریخ ۱۳۹۷/۰۶/۱۹ توسط کمیته تخصصی زیر جهت اخذ مدرک رساله دکتری ارزیابی گردید و با درجه ...بسیار خوب... مورد پذیرش قرار گرفت.

| اساتید راهنما | امضاء | اساتید مشاور | امضاء |
|---|---|---|---|
| نام و نام خانوادگی:<br>آقای دکتر حمید حسن‌پور | | نام و نام خانوادگی:<br>خانم دکتر هدی مشایخی | |

| اساتید داور | امضاء | نماینده تحصیلات تکمیلی | امضاء |
|---|---|---|---|
| نام و نام خانوادگی:<br>آقای دکتر بهروز مینایی | | نام و نام خانوادگی:<br>آقای دکتر منصور فاتح | |
| نام و نام خانوادگی:<br>آقای دکتر مرتضی زاهدی | | | |
| نام و نام خانوادگی:<br>آقای دکتر علیرضا احمدی فرد | | | |

وزارت علوم، تحقیقات و فناوری
دانشکده فناوری اطلاعات و کامپیوتر

وزارت علوم، تحقیقات و فناوری
دانشکده فناوری اطلاعات و کامپیوتر

# تعهدنامه

اینجانب **باقر رحیم‌پورکامی**دانشجوی دوره **دکتری مهندسی هوش مصنوعی دانشکده مهندسی کامپیوتر** دانشگاه صنعتی شاهرود نویسنده پایان نامه" **ارائه سیستم توصیه‌گر خبر مبتنی بر تمایلات کاربر**" تحت راهنمائی آقای **دکتر حمید حسن‌پور**متعهد می‌شوم :

- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش‌های محققان دیگر به مرجع مورد استفاده استناد شده است.
- مطالب مندرج در پایان نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است.
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود است و مقالات مستخرج با نام **دانشگاه صنعتی شاهرود** و یا **Shahrood University of Technology** به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان نامه رعایت می‌گردد.
- در کلیه مراحل انجام این پایان نامه، در مواردی که از موجود زنده( یا بافت‌های آن‌ها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است.
- در کلیه مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری، ضوابط و اصول اخلاق انسانی رعایت شده است.

**تاریخ :**

**امضاء دانشجو**

**مالکیت نتایج و حق نشر**

- کلیه حقوق معنوی این اثر و محصولات آن(مقالات، مستخرج، کتاب، برنامه‌های رایانه ای، نرم‌افزارها و تجهیزات ساخته شده) متعلق به دانشگاه صنعتی شاهرود می‌باشد. این مطلب باید یه نحو مقتضی در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات نتایج موجود در پایان نامه بدون ذکر مرجع مجاز نمی‌باشد.

## چکیده

سیستم‌های توصیه‌گر به عنوان یک فناوری مهم جهت کمک به کاربران در انتخاب منابع مناسب، اخیرا رواج پیدا کرده و به سرعت در حال توسعه می‌باشند. این سیستم‌ها با بررسی تعاملات گذشته کاربران و شناسایی سلیقه آنها، یک محیط شخصی‌سازی شده جهت انتخاب منابع را فراهم می‌کنند. مدلسازی رفتار کاربر و مکانیسم ارائه توصیه ازمسائل اساسی در توسعه سیستم‌های توصیه‌گر می‌باشند.

روش‌های موجود، غالبا سلیقه کاربر را به عنوان یک فاکتور پنهان در مدلسازی رفتار کاربر بکار می‌گیرند. پارامتر سلیقه از پروفایل کاربر استنتاج می‌شود. هر سلیقه کاربر بیان کننده مجموعه‌ای از منابع با ویژگی‌های مشترک (مشابه) می‌باشد که در گذشته توسط آن کاربر انتخاب شده است. عدم توجه به پویایی سلایق کاربر و عدم اولویت‌بندی بین سلایق مختلف یک کاربر در زمان‌های مختلف از چالش‌های اساسی روش‌های موجود می‌باشد.

در این پژوهش، برای مدلسازی رفتار کاربر علاوه بر مولفه سلیقه، از مولفه علاقه (تمایل) نیز استفاده شده است که بیان‌گر اهمیت هر یک از سلیقه‌های کاربر در زمان معین است. تعداد، زمان، و امتیاز منابع انتخاب شده از عوامل اساسی در محاسبه تمایل کاربر می‌باشند. در روش پیشنهادی، ابتدا با بکارگیری پروفایل کاربر در یک چارچوب شبکه بیزین به نام فرآیند دریکله، رفتار هر کاربر با استفاده از زوج مولفه <**سلیقه ، تمایل**>، مدل می‌شود. سپس، براساس مدل رفتاری ایجاد شده، کاربران مشابه (همسایه)، گردآوری می‌شوند. در پایان، منابع انتخاب شده توسط کاربران مشابه به کاربر متقاضی پیشنهاد می‌شود. پویایی تعداد سلایق و سازگاری تمایل کاربر از مزیت‌های اساسی روش پیشنهادی می‌باشند. همچنین در مدل پیشنهادی، تنها یک ابرپارامتر (عددی) استفاده شده که مقدار آن نیز بصورت ابتکاری محاسبه می‌شود. در این رساله، یک سیستم توصیه‌گر خبر به عنوان مطالعه موردی برای پیاده‌سازی مدل پیشنهادی انتخاب شده است.

برای ارزیابی مدل پیشنهادی از داده‌های جمع‌آوری شده از بستر شبکه توئیتر (کانال‌های خبری) و نیز پایگاه داده MovieLens استفاده شده است. کارائی مدل پیشنهادی نیز با روش‌های معروف نظیر TimeSVD, Content-based, non-Negative Matrix Factorization, Item-Item Collaborative Filtering مقایسه شده است. نتایج حاصل از ارزیابی و مقایسه، کارائی مدل پیشنهادی را تایید می‌کند.



**واژه‌های کلیدی:** سیستم‌های توصیه‌گر، مدلسازی کاربر، الگوریتم توصیه‌گر ترکیبی، سیستم توصیه‌گر خبر.

## مقالات مستخرج از رساله

| | |
|---|---|
| **1** | *ISI Journal.*<br>Bagher, Rahimpour Cami, Hamid Hassanpour, and Hoda Mashayekhi.<br>**"User trends modeling for a content-based recommender system."**<br>*Expert Systems with Applications* 87 (2017): 209-219. |
| **2** | *ISI Journal.*<br>Bagher, Rahimpour Cami, Hamid Hassanpour, and Hoda Mashayekhi.<br>**"User Preferences Modeling using Dirichlet Process Mixture Model for a Content-Based Recommender System."**<br>*Knowledge-Based Systems. Accepted, Aug 26 2018.* |
| **3** | *ISI Journal.*<br>Bagher, Rahimpour Cami, Hamid Hassanpour, and Hoda Mashayekhi.<br>"**A Preference-based User Similarity to Construct a Collaborative Recommender System.**"<br>*Information Sciences (submitted).* |
| **4** | Cami, Bagher Rahimpour, Hamid Hassanpour, and Hoda Mashayekhi.<br>**"A content-based movie recommender system based on temporal user preferences."**<br>In *Intelligent Systems and Signal Processing (ICSPIS), 2017 3rd Iranian Conference on*, pp. 121-125. IEEE, 2017. |
| **5** | Cami, Bagher Rahimpour, Hamid Hassanpour, and Hoda Mashayekhi.<br>**"Inferring User Preference Using Distance Dependent Chinese Restaurant Process and Weighted Distribution."**<br>In *Artificial Intelligence and Signal Processing Conference (AISP), 2017*. IEEE, 2017. |

## فهرست مطالب


فصل اول: سیستم‌های توصیه‌گر برای دسترسی منابع در بستر وب ..... ۱
۱-۱- مقدمه ..... ۲
۱-۲- معرفی سیستم توصیه‌گر ..... ۳
۱-۲-۱- مشکلات پیاده‌سازی سیستم‌های توصیه‌گر ..... ۶
۱-۲-۲- رویکرد الگوریتمی در طراحی سیستم توصیه‌گر ..... ۸
۱-۲-۳- رویکرد مدلسازی رفتار کاربر در طراحی سیستم توصیه‌گر ..... ۹
۱-۳- بیان مسئله ..... ۱۰
۱-۴- نوآوری و دستاوردها ..... ۱۲
۱-۵- مفروضات ..... ۱۵
۱-۶-ارزیابی ..... ۱۶
۱-۷- ساختار رساله ..... ۱۷
فصل دوم : بررسی روش‌های موجود برای توسعه سیستم‌های توصیه‌گر ..... ۱۹
۲-۱- مقدمه ..... ۲۰
۲-۲- مروری بر رویکرد الگوریتمی در توسعه سیستم‌های توصیه‌گر ..... ۲۰
۲-۳- مروری بر رویکرد مدلسازی رفتار کاربر در توسعه سیستم‌های توصیه‌گر ..... ۴۷
۲-۴- ارزیابی سیستم‌های توصیه‌گر ..... ۴۸
۲-۵- جمع‌بندی ..... ۵۷
فصل سوم : روش پیشنهادی مدلسازی رفتار کاربر برای توسعه یک سیستم توصیه‌گر ..... ۵۹
۳-۱- مقدمه ..... ۶۰
۳-۲- تحلیل مولفه‌های موثر بر رفتار کاربر در انتخاب منابع (نوآوری) ..... ۶۱
۳-۳- فرآیند مدلسازی دریکله ..... ۶۷
۳-۴- ارائه مدل پیشنهادی کاربر (نوآوری) ..... ۶۹
۳-۵- بکارگیری مدل پیشنهادی جهت ارائه توصیه (نوآوری) ..... ۷۵
۳-۵-۱- ایجاد سیستم توصیه‌گر بر اساس مولفه‌های سلیقه و علاقه (نوآوری) ..... ۷۵
۳-۵-۲- ایجاد سیستم توصیه‌گر ترکیبی (نوآوری) ..... ۷۶
۳-۶- جمع‌بندی ..... ۷۸
فصل چهارم : آزمایش‌ها و تحلیل نتایج ..... ۸۱
۴-۱- مقدمه ..... ۸۲

۴-۲- پایگاه داده و محیط پیاده‌سازی ..... ۸۲
۴-۳- معیارهای ارزیابی و روش‌های مورد مقایسه‌ای ..... ۸۳
۴-۴- آزمایش‌ها و نتایج ..... ۸۳
۴-۴-۱- ارزیابی صحت مدل پیشنهادی ..... ۸۴
۴-۴-۲- بررسی پارامتر توزیع سلایق کاربر ..... ۹۰
۴-۴-۳- بررسی تاثیر مولفه تمایل بر معیار شباهت در پالایش مشارکتی ..... ۹۲
۴-۵- جمع‌بندی ..... ۹۳
فصل پنجم : جمع‌بندی و کارهای آینده ..... ۹۵
۵-۱- جمع‌بندی پژوهش انجام شده در رساله ..... ۹۶
۵-۲- کارهای آینده ..... ۹۸
فهرست واژگان ..... ۹۷
مراجع ..... ۱۰۳

## فهرست شکل‌ها


شکل ۱-۱: قالب کلی یک سیستم توصیه گر که به صورت سه تایی (ورودی- پردازش- خروجی) مدل شده است. ---------- ۵

شکل ۲-۱: شمای کلی سیستمهای توصیه گر مبتنی بر فیلترینگ مشارکتی. ---------- ۲۲

شکل ۲-۲: مراحل پالایش مشارکتی مبتنی بر منبع را نشان میدهد. ---------- ۲۶

شکل ۲-۳: قسمتی از ارتباط بین کاربران در شبکه اجتماعی که به کمک شبکه بیز مدل شدهاست را نشان میدهد. ---------- ۲۹

شکل۲-۴: نحوه محاسبه ماتریسهای U و V را نشان میدهد. تصویر a، ماتریس بازخورد اولیه و تصویر b، ماتریس پیشبینی شده را نشان میدهند [۷۰]. ---------- ۳۲

شکل۲-۵: فرایند انجام توصیه در الگوریتم مبتنی بر محتوا را نشان میدهد. ---------- ۳۴

شکل۲-۶: نمایش فضای سه بعدی از ماتریس بازخورد با درنظر گرفتن زمان به عنوان یک مولفه در زمینه تعامل [۲۳]. ---------- ۴۱

شکل ۳-۱: نتیجه آزمایش مربوط به تاثیر مولفههای سلیقه و تمایل در انتخاب منابع را نشان میدهد. محور افقی، بیان کننده شناسه سلیقه کاربر میباشد. ---------- ۶۴

شکل ۳-۲: نمایش آزمایش مربوط به میزان تمایل کاربر به انتخاب منابع از هر سلیقه در دو بازه زمانی متوالی. محور افقی، بیان کننده شناسه سلیقه کاربر میباشد. ---------- ۶۵

شکل ۳-۳: افرازهای فضای مسئله جهت نمایش توزیع دریکله. ---------- ۶۷

شکل ۳-۴: مدل گرافیکی احتمالاتی برای نمایش مدل پیشنهادی رفتار کاربر. ---------- ۷۰

شکل ۳-۵: فرآیند محاسبه پارامترها و ایجاد مدل. ---------- ۷۴

شکل ۳-۶: فرآیند بکارگیری مدل کاربر و تهیه لیست پیشنهادی. ---------- ۷۶

شکل۳-۷: الگوریتم ارائه توصیه با رویکرد ترکیبی. ---------- ۷۸

شکل ۳-۸: معماری سیستم توصیهگر ترکیبی پیشنهادی. ---------- ۷۸

شکل ۴-۱: بررسی تاثیر مدل پیشنهادی در کارائی سیستم توصیهگر و مقایسه دو روش مدلسازی "مبتنی بر سلیقه" و مبتنی بر زوج مولفه "سلیقه و تمایل". ---------- ۸۷

شکل۴-۲: ارزیابی سازگاری مدل پیشنهادی در مواجه با مشاهدات جدید. ---------- ۸۹

شکل ۴-۳: نمایش همبستگی مقدار اولیه پارامتر توزیع سلایق با مقدار بهینه آن. ---------- ۹۱

## فهرست جدول‌ها

جدول ۲-۱: دستهبندی الگوریتمهای موجود برای سیستم توصیهگر براساس تکنیکها و داده ورودی. - ۲۱
جدول۲-۲: مقایسه روشهای مختلف ارزیابی سیستمهای توصیهگر. ----------۵۰
جدول۲-۳: حالت های مختلف مفید بودن /نبودن منبع و انتخاب هریک از منابع. ----------۵۲
جدول ۳-۱: نمادهای استفاده شده برای بیان مدل پیشنهادی. ----------۶۶
جدول۴-۱: مشخصات آماری پایگاه داده های استفاده شده در آزمایشات. ---------- ۸۲
جدول ۴-۲: نتایج اجرای روش پیشنهادی و الگوریتم LOGO. ---------- ۸۵
جدول ۴-۳: مقایسه نتایج اجرای روش پیشنهادی و ابزارهای معروف. ---------- ۸۵
جدول ۴-۴: مقایسه عملکرد روش پیشنهادی و روشهای موجود در مواجه با مسئله تنکی داده. ---- ۸۸
جدول ۴-۵: مقایسه عملکرد روش پیشنهادی با روش موجود مبتنی بر پالایش مشارکتی. ---------۹۲

## فهرست اختصارات

**C**

CF …..... Collaborative Filtering

**D**

ddCRP …..... distance dependent Chienece Restaurant Process

DPMM…..... Dirichlet Process Mixture Model

**E**

EM …..... Expectation Maximization

**I**

IBP …..... Indian buffet process

**L**

LDA …..... Latent Dirichlet Allocation

**N**

nDCG…..... normalized Discounted Cumulative Gain

NMF …..... Non-negative matrix factorization

**P**

PLSA …..... Probabilistic Latent Semantic Analysis

**S**

SVD …..... Singular Value Decomposition

# فصل اول

# سیستم‌های توصیه‌گر برای دسترسی منابع در بستر وب

## ۱-۱- مقدمه

سیستم‌های خدماتی تحت وب نظیر تجارت الکترونیک، آموزش الکترونیک، و اخبار حجم انبوهی از منابع را فراهم آورده‌اند. دستیابی سریع به منابع مناسب یکی از چالش‌های مهم در مواجهه با حجم انبوه منابع در این حوزه می‌باشد. موتورهای جستجو سنتی نظیر Google و Yahoo، به عنوان یکی از ابزارهای موجود جهت کمک به دستیابی منابع مورد استفاده قرار می‌گیرند. عدم توجه به سلیقه کاربر از مشکلات عمده این موتورهای جستجو می‌باشد. عدم توجه به سلیقه کاربر سبب می‌شود تا این موتورهای جستجو نتوانند منابع مورد علاقه کاربر را به وی پیشنهای نمایند و کاربر مجبور به صرف زمان زیاد برای گشت و گذار و یافتن منابع مورد علاقه می‌باشد. همچنین سیستم‌های تجاری امکاناتی نظیر پیوندهای **تازه‌ترین‌ها** و **پرفروش‌ترین‌ها** را برای کمک به کاربران جهت یافتن منابع مناسب بکار می‌گیرند [۱]. هر چند وجود این امکانات می‌تواند مفید باشد اما مشکل اساسی این است که مطابق سلیقه (نیاز) هر کاربر، اطلاعات فراهم نمی‌شود.

سیستم‌های توصیه‌گر به عنوان یکی از فناوری‌های موفق برای کمک به کاربران و فراهم‌کنندگان منابع درحال توسعه می‌باشند. سیستم‌های توصیه‌گر با بکارگیری تکنیک‌ها و ابزارهای هوش مصنوعی، محیط شخصی‌سازی شده‌ای را برای کاربران فراهم می‌کنند. این محیط ضمن اینکه منابع متناسب با سلیقه کاربران را ارائه می‌کند، زمینه بازاریابی و ارائه منابع جدید توسط فراهم‌کنندگان منابع را نیز امکان‌پذیر می‌سازد. به عبارتی، هدف اصلی سیستم‌های توصیه‌گر شخصی‌سازی محیط تعاملی و در نتیجه ارائه منابع متناسب با سلیقه (نیاز) کاربر می‌باشد. در حال حاضر از این فناوری در حوزه‌های مختلف تجاری نظیر Amazon (فروشگاه اینترنتی)، Netflix (خدمات ارائه فیلم)، GoogleNews (خدمات ارائه خبر) استفاده می‌شود [۱-۳]. در ادامه ضمن معرفی سیستم توصیه‌گر، بیان مسئله، نوآوری، مفروضات، و چگونگی ارزیابی نیز ارائه می‌شوند.

## ۱-۲- معرفی سیستم توصیه‌گر

سیستم‌های توصیه‌گر به عنوان یک ابزار مفید جهت ایجاد محیط شخصی‌سازی شده در بستر وب در حال توسعه می‌باشند. سیستم‌های توصیه‌گر با بکارگیری تراکنش‌های گذشته کاربران در مکانیسم-ها و تکنیک‌های یادگیری ماشین نظیر مدل‌های احتمالاتی و خوشه‌بندی، ضمن ایجاد تخمین مناسب از رفتار آینده کاربران، سعی می‌کنند منابع مورد علاقه کاربران را شناسایی و پیشنهاد نمایند [۴]. هدف اصلی این سیستم‌ها کمک به کاربران جهت یافتن منابع متناسب با سلیقه از بین تعداد زیادی از منابع موجود در بستر وب می‌باشد [۵]. نحوه بکارگیری تعاملات کاربران جهت استنتاج سلایق آنها، از چالش‌های اساسی در توسعه سیستم‌های توصیه‌گر می‌باشد. در حوزه توسعه سیستم‌های توصیه‌گر، اطلاعات حاصل از تعاملات کاربران، به عنوان پروفایل کاربران شناخته می‌شود. پروفایل کاربران شامل اطلاعاتی نظیر ویژگی کاربران، ویژگی منابع، بازخورد کاربران، اطلاعات زمینه تعامل می‌باشد [۶]. با توجه به شکل ۱-۱، سیستم‌های توصیه‌گر را می‌توان با سه مولفه ورودی، پردازشی، و خروجی مدل نمود [۷]. این مولفه‌ها در ادامه توضیح داده می‌شوند:

**مولفه ورودی**. این مولفه که عموما به عنوان پروفایل کاربر شناخته می‌شود، شامل بازخوردها، ویژگی کاربر، ویژگی منبع، شبکه اجتماعی، پایگاه دانش، و زمینه تعامل می‌باشد. بازخوردها نشان دهنده میزان مطلوبیت (کمی یا کیفی) منابع از نظر کاربران می‌باشند. این بازخوردها ممکن است بصورت صریح[1] یا ضمنی[2]، توسط کاربران ثبت شوند [۸, ۹]. کاربر دارای ویژگی‌های ساختاری نظیر نام، سن؛ و یا ویژگی‌های پویا مانند سلیقه، زمان و فعالیت‌ها می‌باشد [10]. منبع شامل هر سرویس یا محصولی ( نظیر کتاب، فیلم، و خبر) می‌باشد که در محیط وب ارائه می‌گردد. منبع نیز دارای ویژگی-های ساختاری (نظیر آنتالوژی) و یا غیرساختاری (دنباله‌ای از کلمات) می‌باشد [۱۰]. دانش اجتماعی از طریق ارتباطات کاربران یا منابع در بستر شبکه‌های اجتماعی استخراج می‌شود. پایگاه دانش شامل

[1] - Explicit
[2] - Implicit

قوانین و محدودیت‌های مربوط به یک محیط تعامل می‌باشد. زمینه تعامل نیز شامل اطلاعاتی نظیر زمان، مکان، و اعتماد[1] در هنگام تعامل بین کاربر و سیستم توصیه‌گر در راستای انتخاب یک منبع مناسب می‌باشد.

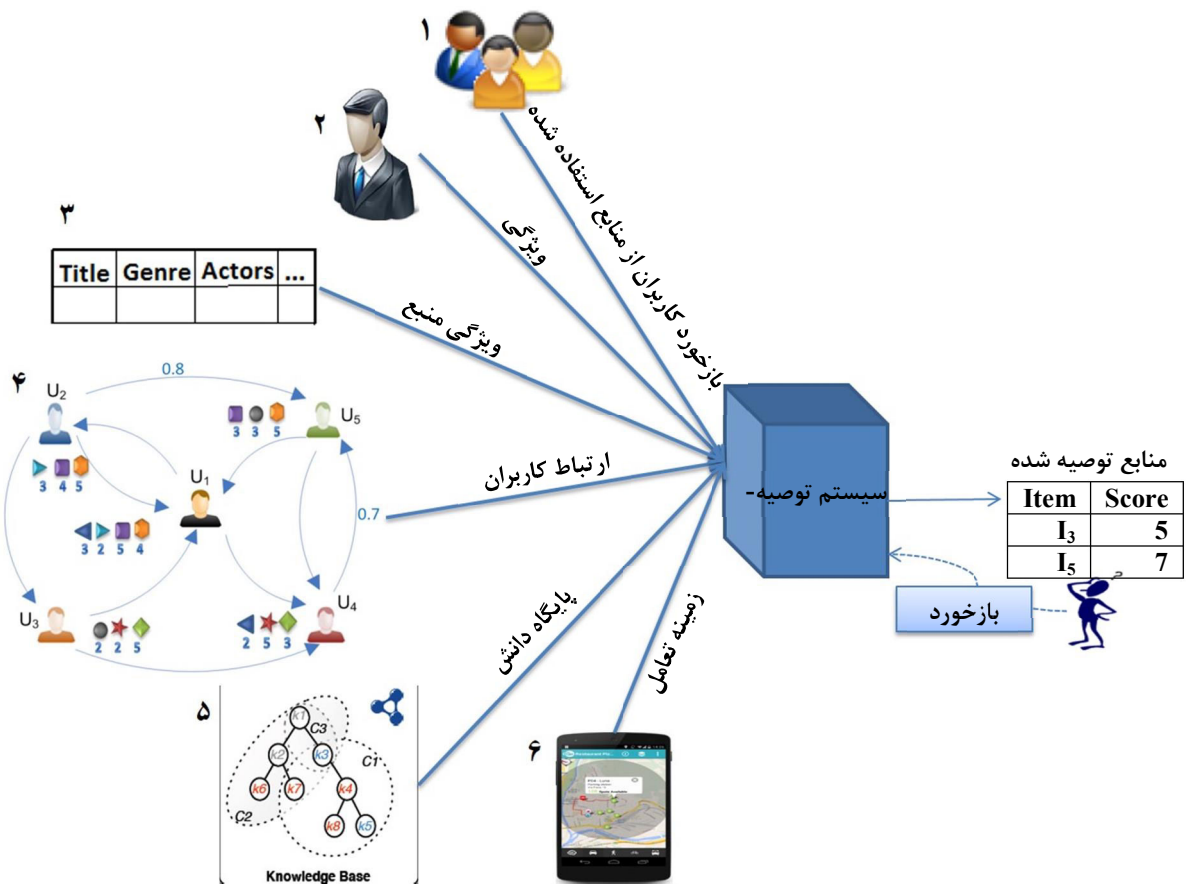
**مولفه پردازش**. این مولفه شامل مکانیسم‌ها و الگوریتم‌های ارائه توصیه می‌باشد که با استفاده از داده‌های ورودی، توصیه مناسب را انجام می‌دهد. انتخاب مکانیسم مناسب جهت ارائه توصیه به محیط تعاملی و داده‌های موجود بستگی دارد. پالایش مشارکتی[2] و پالایش محتوایی[3] به عنوان دو رویکرد اساسی در مولفه پردازشی مورد استفاده قرار می‌گیرند [۳, ۶]. تشریح مکانیسم‌ها و الگوریتم‌های موجود در فصل دوم آمده‌است.

**مولفه خروجی**. این مولفه نیز نحوه ارائه خروجی را نشان می‌دهد. خروجی عموما به دو صورت پیش‌بینی محبوبیت (امتیاز) منابع و ارائه لیست مرتب شده از منابع ارائه می‌شود [۳]. در مولفه خروجی ممکن است بازخورد کاربر نیز دریافت و ذخیره گردد.

سیستم‌های تجاری مختلف با بکارگیری فناوری سیستم‌های توصیه‌گر در جهت شخصی‌سازی محیط تعامل سعی در کمک به کاربران و جذب بیشتر مشتریان دارند. کاربرد سیستم‌های توصیه‌گر را می‌توان در حوزه‌های مختلف نظیر فروشگاه اینترنتی جهت پیشنهاد کالای مناسب به مشتریان (فروشگاه الکترونیکی)[4] [۱]، کتابخانه‌های دیجیتال جهت پیشنهاد کتاب مناسب (آموزش [۱۱]، پزشکی[6] [۱۲] به منظور پیشنهاد پزشک متخصص یا مراقبت درمانی با توجه به نوع بیماری افراد، و در شبکه‌های اجتماعی جهت توصیه افراد[7] مشاهده نمود [۱۳, ۱۴].

---

[1] - Trust
[2] - Collaborative Filtering (CF)
[3] - Content-based Filtering
[4] - http://www.amazon.com
[5] - http://www.atutor.ca
[6] - http://www.patientslikeme.com
[7] - http://www.epinions.com




شکل ۱-۱: قالب کلی یک سیستم توصیه‌گر که به صورت سه تایی (ورودی- پردازش- خروجی) مدل شده است.

در ایجاد و توسعه سیستم‌های توصیه‌گر، دو مسئله اساسی مورد توجه می‌باشد. مسئله اول، رویکرد الگوریتمی برای ایجاد سیستم توصیه‌گر می‌باشد که به نحوه بکارگیری تکنیک‌ها و روش‌ها برای ایجاد یک سیستم توصیه‌گر پرداخته می‌شود. مسئله دوم، رویکرد مدلسازی کاربر می‌باشد که پارامترهای تشکیل دهنده پروفایل کاربر، نحوه استنتاج، و بکارگیری آن بررسی می‌شود. این رویکردها سعی در غلبه بر مشکلاتی دارند که سیستم‌های توصیه‌گر با آن مواجه می‌باشند. در ادامه، ضمن بیان این مشکلات، هر یک از این رویکردها تشریح می‌شوند.

### ۱-۲-۱- مشکلات پیاده‌سازی سیستم‌های توصیه‌گر

با آغاز توسعه سیستم‌های توصیه‌گر از سال ۱۹۹۰ تلاش زیادی برای ارائه مکانیسم مناسب برای ایجاد سیستم‌های توصیه‌گر انجام شده‌است [۱۵]. هر یک از الگوریتم‌های ارائه شده سعی در بهبود قسمتی از مشکلاتی دارند که سیستم‌های توصیه‌گر با آنها مواجه می‌باشند. در ادامه مواردی از این مشکلات بیان شده‌است.

- **تنکی داده**[1]: با وجود اینکه منابع و کاربران زیادی در محیط وب تعامل دارند و حجم عظیمی از داده‌ها را فراهم می‌آورند، اما ممکن است یک کاربر تعداد معدودی از منابع را انتخاب نماید و یا برای تعداد معدودی از منابع انتخاب شده بازخورد ثبت نماید. با این ترتیب با توجه به تنک بودن ماتریس بازخورد، ساخت ویژگی (ایجاد مدل) مناسب از کاربر و منبع کار مشکلی می‌باشد.
- **تغییر سلیقه کاربر**[2]: سلیقه کاربر معمولا در طول زمان و در شرایط مختلف همراه با تغییر نیاز کاربر در حال تغییر می‌باشد. از اینرو الگوریتم ارائه شده برای سیستم‌های توصیه‌گر باید بتواند نسبت به تغییر سلیقه کاربر حساس باشد و در کوتاه‌ترین زمان ممکن، مدل سلیقه کاربر را به روز نماید و براساس نیاز جدید پیشنهاد منبع را انجام دهد.
- **غیر قابل پیش بینی بودن محتوای برخی از منابع**[3]: برخی از منابع به خاطر ماهیت خود، قابل پیش-بینی نمی‌باشند. به عنوان مثال دو موزیک که توسط یک خواننده منتشر شده‌اند، ممکن است به سبب محتوا از نظر کاربران مختلف، استقبال مختلفی داشته باشند. بنابراین، صرفا ویژگی خواننده نمی‌تواند برای تصمیم‌گیری مناسب باشد.
- **نحوه دریافت بازخورد کاربر**[4]: بازخورد کاربر یکی از مولفه‌های مهم برای سیستم‌های توصیه‌گر جهت مدل‌کردن کاربر و منابع می‌باشد. نحوه دریافت بازخورد کاربر در مشارکت کاربر برای ارائه بازخورد

---

[1] - Sparsity(Lack of Data)
[2] - Changing User Interest
[3] - Unpredictable Items
[4] - Implicit/Explicit Rating

مهم می‌باشد. در برخی روش‌ها از تعامل مستقیم کاربر برای دریافت بازخورد استفاده می‌شود. در برخی روش‌ها از بازخورد غیر مستقیم کاربر مانند تعداد بازدیدها، افزودن منبع به لینک‌های مورد علاقه در مرورگر، افزودن منبع به سبد خرید، دانلود و مدت زمان بازبینی استفاده می‌شود. این روش هر چند کاربر را درگیر ارائه بازخورد نمی‌کند اما میزان خطا در این روش نسبت به بازخورد صریح بیشتر است [۱۶].

- **توسعه‌پذیری و دقت**: یکی از مشکلات اساسی که سیستم‌های توصیه‌گر با آن روبرو هستند، قابلیت بکارگیری آنها در سیستم‌های تجاری می‌باشد. حجم زیاد داده (تعداد زیاد کاربران، منابع، و تعاملات)، مستلزم وجود الگوریتم‌های کارا (با پیچیدگی زمانی مناسب) جهت ارائه پیشنهاد در زمان مناسب می‌باشد. همچنین الگوریتم پیشنهادی باید بتواند تخمین مناسبی از سلیقه ارائه نماید و براساس آن پیش‌بینی درستی از رفتار کاربر در انتخاب منابع داشته باشد.
- **منبع جدید/ کاربر جدید (شروع سرد)**[1]: عملکرد سیستم‌های توصیه‌گر براساس داده‌های موجود از سابقه تراکنش‌های کاربر (نظیر ویژگی کاربر و منبع) می‌باشد. از اینرو در مورد منابع و کاربرانی که برای اولین بار با سیستم تعامل دارند (یا تعداد تعاملات آنها بسیار محدود باشد)، سیستم نمی‌تواند توصیه مناسبی را ارائه نماید.

با توجه به مشکلات بیان شده و راهکارهای موجود، عموما سیستم‌های توصیه‌گر به عنوان سیستم‌های تصمیم‌یار برای کمک به کاربران در بستر وب مورد استفاده قرار می‌گیرند و به عنوان جایگزین انسان در تصمیم‌گیری برای انتخاب منابع نمی باشند. با این وجود، تجربه‌های موفق نظیر Amazon[2] و Netflix[3] از بکارگیری این فناوری در سیستم‌های تجاری وجود دارند که کارآمدی سیستم‌های توصیه‌گر را نشان می‌دهند [۱۷, ۱۸]. در ادامه رویکردهای اساسی برای توسعه سیستم‌های توصیه‌گر تشریح می‌شوند.

---

[1] - Cold Start
[2] - https://www.amazon.com
[3] - https://www.netflix.com

### ۱-۲-۲- رویکرد الگوریتمی در طراحی سیستم توصیه‌گر

الگوریتم‌ها و تکنیک‌های (یادگیری ماشین) مختلفی در ایجاد سیستم‌های توصیه‌گر مورد استفاده قرار می‌گیرند. این الگوریتم‌ها بر نحوه بکارگیری پروفایل کاربر در جهت تحلیل و استخراج دانش برای پیشنهاد منابع تمرکز دارند. رویکردهای مختلفی برای بکارگیری پروفایل کاربران و ایجاد یک سیستم توصیه‌گر وجود دارد. این رویکردها را می‌توان به پنج دسته تقسیم نمود که شامل پالایش مشارکتی، پالایش محتوایی، توصیه مبتنی‌بر زمینه تعامل، توصیه مبتنی‌بر پایگاه دانش، و رویکردهای ترکیبی می‌باشد [۶, ۱۹]. در رویکرد پالایش مشارکتی، یک چهارچوب معین برای بکارگیری پروفایل کاربران و اعمال یک معیار شباهت برای استخراج گروه‌هایی از کاربران مشابه ایجاد می‌شود [۱, ۲۰, ۲۱]. ایده اصلی رویکرد پالایش مشارکتی این است که کاربران مشابه (دارای سلیقه مشترک)، منابع مشابه را انتخاب می‌کنند. در این رویکرد، عموما پروفایل کاربران شامل بازخورد ( نظیر امتیازدهی و تقه) کاربران نسبت به منابع انتخاب شده می‌باشد. در رویکرد پالایش محتوایی، پروفایل کاربران عموما شامل ویژگی منابعی می‌باشد که هر کاربر در طول زمان انتخاب نموده‌است. بدین ترتیب، با استفاده از تکنیک‌های مقایسه‌ای مناسب، منابع مشابه با پروفایل کاربر متقاضی، انتخاب و پیشنهاد می‌شود [۱۰, ۲۲]. در رویکرد مبتنی‌بر زمینه تعامل، اطلاعات زمینه تعامل نظیر زمان، مکان، و شبکه اجتماعی در الگوریتم ارائه توصیه استفاده می‌شوند [۲۳]. در رویکرد مبتنی‌بر پایگاه دانش، نیازمندی‌های کاربران و مشخصه‌های منابع در یک پایگاه دانش گردآوری می‌شود. سپس با اعمال تکنیک‌های استنتاجی و استدلالی بر پایگاه دانش، منابعی انتخاب می‌شوند که به نیازمندی‌های کاربر نزدیک‌تر باشند [۲۴]. با توجه به حوزه‌های کاربردی و پارامترهایی که برای تشکیل پروفایل کاربر وجود دارد، ممکن است ترکیب روش‌های فوق بتواند نتایج بهتری را تولید نماید. از اینرو در رویکرد ترکیبی، براساس پروفایل کاربر (حوزه کاربرد) دو یا چند روش فوق با هم ادغام می‌شوند و یک الگوریتم ترکیبی برای پیشنهاد منابع ایجاد شود [۲۵]. با توجه به رویکردهای بیان شده، روش‌های متعددی برای ایجاد سیستم‌های

توصیه‌گر ایجاد شده‌اند. این روش‌ها در فصل دوم (بررسی روش‌های موجود برای سیستم‌های توصیه-گر)، تشریح می‌شوند.

### ۱-۲-۳- رویکرد مدلسازی رفتار کاربر در طراحی سیستم توصیه‌گر

پروفایل کاربران اطلاعاتی نظیر ویژگی کاربران، ویژگی منابع، و بازخورد (امتیازدهی یا کلیک) کاربران را نگهداری می‌کند. این پروفایل برای استنتاج سلایق کاربران و پیش‌بینی رفتار آتی کاربران مورد استفاده قرار می‌گیرد [۲۶]. رویکرد مدلسازی بر انتخاب پارامترهای موثر و نحوه ارائه آن جهت بکارگیری در الگوریتم توصیه تمرکز دارد. مدلسازی پروفایل کاربر (یا به عبارتی مدلسازی کاربر) به دو روش مدلسازی ساختاری [۲۷, ۲۸] و مدلسازی رفتاری [۲۹, ۳۰] انجام می‌شود.

در مدلسازی **ساختاری**، ویژگی‌هایی از پروفایل کاربر (کاربران) استخراج و یک مدل تجریدی از آن ایجاد می‌شود [۳۱]. به عنوان مثال در محیطی که کاربر به منابع انتخاب شده امتیازدهی می‌کند (نظیر سامانه دانلود فیلم)، پروفایل کاربر بصورت بردار $< \text{user\_id}, \text{item\_id}, \text{rate}, \text{timestamp} >$ در الگوریتم توصیه بکار گرفته می‌شود. رویکرد پالایش مشارکتی مبتنی‌بر همسایه [۱, ۹] و فاکتورهای پنهان [۳۲, ۳۳] از این مدل استفاده می‌کنند. همچنین در محیط‌هایی که محتوای منبع مشخص باشد (نظیر سامانه‌های خبری)، پروفایل کاربر با استفاده از بردار کلمات کلیدی مدل می‌شود. در رویکرد پالایش محتوایی، از این دسته از مدلسازی‌ها استفاده می‌شود که با مقایسه پروفایل کاربر با بردار کلمات کلیدی هر منبع، پیشنهاد منبع انجام می‌شود [۳۴].

در مدلسازی **رفتاری** از چارچوب احتمالاتی مانند شبکه بیزین و مدل مارکوف برای مدلسازی پروفایل استفاده می‌شود [۳۰, ۳۵]. در این رویکرد، با شناسایی پارامترهایی (قابل مشاهده یا مخفی) که رفتار کاربر را منعکس می‌کنند، یک مدل گرافیکی احتمالاتی (Probabilistic Graphical Model) ایجاد می‌شود. در این مدل، عموما سلیقه کاربر با استفاده از پارامترهای پنهان نشان داده می‌شوند. با استنتاج پارامترها، سلیقه کاربر شناسایی و براساس آن، پیشنهاد منبع انجام می‌شود. در

مدل رفتاری، براساس نحوه انتخاب پارامترها (مبتنی‌بر پروفایل یک کاربر یا تجمعی) علاوه بر سلایق فردی، می‌توان سلایق گروهی را نیز شناسایی نمود.

## ۱-۳-بیان مسئله

صرفنظر از رویکردهای مدلسازی و مکانیسم‌های تولید پیشنهاد منابع، شناسایی و استنتاج سلایق کاربر به عنوان مهمترین مولفه در روش‌های موجود برای توسعه سیستم‌های توصیه‌گر می‌باشد. ایجاد مدل مناسب از رفتار کاربر به عنوان یکی از راهکارهای اساسی در پیش‌بینی رفتار کاربر و بهبود عملکرد سیستم توصیه‌گر می‌باشد. در روش‌های موجود از مولفه **سلیقه** کاربر برای مدل کردن رفتار کاربر استفاده می‌شود [۲۹, ۳۶]. منظور از سلیقه کاربر، گروه‌هایی از منابع مشابه می‌باشند که کاربر آنها را در طول زمان انتخاب کرده‌است. روش‌های موجود، با استخراج سلایق کاربر، منابع متناسب با این سلایق را به کاربر پیشنهاد می‌کنند. به عبارتی یک کاربر با استفاده از بردار سلیقه مدل می‌شود. هر مولفه از این بردار، مجموعه‌ای از منابع مشابه را نشان می‌دهد که کاربر در طول زمان انتخاب نموده‌است. بدین ترتیب با استفاده از بردار سلیقه برای هر کاربر، می‌توان منابع مرتبط با سلیقه کاربر را انتخاب نمود.

از مسائل اساسی در روش‌های موجود، عدم توجه به میزان اهمیت هر سلیقه و عدم پویایی آنها در طول زمان می‌باشد. در این روش‌ها، اولویت‌بندی برای سلیقه‌های یک کاربر وجود ندارد و به تبع آن، منابع پیشنهادی نیز براساس اولویت سلایق مرتب نمی‌شوند. به عنوان مثال در روش‌های مبتنی بر فاکتور پنهان[1] [۳۳, ۳۷]، سلیقه‌ها در یک بازه زمانی معین و براساس پروفایل مجموعه‌ای از کاربران استخراج می‌شوند. در این دسته از روش‌ها، به ازای همه کاربران تعداد سلیقه‌ها ثابت می‌باشد و با انجام تعاملات جدید، اهمیت سلیقه‌ها و تعداد آنها تغییر نمی‌کند. همچنین این دسته از روش‌ها با مشکل منبع جدید نیز مواجه می‌باشند و نمی‌توانند منبع جدید را به کاربران پیشنهاد نمایند.

[1] - Latent Factor Models

هدف اصلی در این رساله، تعریف و استفاده از مولفه **تمایل**[1] کاربر علاوه بر مولفه سلیقه برای مدلسازی رفتار کاربر می‌باشد [۳۸]. بدین ترتیب رفتار کاربر با استفاده از زوج مولفه <**سلیقه ، تمایل**> مدل می‌شود. منظور از تمایل کاربر، میزان علاقه کاربر در انتخاب منابع از هر سلیقه در زمان معین می‌باشد. به عبارتی مولفه تمایل، میزان اهمیت یک سلیقه در یک زمان معین را نشان می‌دهد. به عنوان مثال در حوزه خبر، یک کاربر اخبار مختلفی را در طول زمان مطالعه می‌کند. با دسته‌بندی اخباری که این کاربر در طول زمان مطالعه نموده است، مجموعه اخبار مشابه در کنار هم قرار می-گیرند. این مجموعه‌های خبری می‌توانند شامل اخبار سیاسی، فرهنگی، و ورزشی باشند. بدین ترتیب هر یک از این مجموعه‌ها، سلایق کاربر را نشان می‌دهند. به عبارتی مجموعه اخبار سیاسی، فرهنگی و ورزشی به ترتیب سلایق سیاسی، فرهنگی، و ورزشی را نشان می‌دهند. اما از آنجاییکه این کاربر خبر اقتصادی را مطالعه ننموده است، بنابراین سلیقه اقتصادی ندارد. در این مثال، مولفه تمایل، میزان علاقه کاربر به هر یک از این سلیقه‌ها در یک زمان معین را نشان می‌دهد. به عبارتی علاقه کاربر به انتخاب خبر از هر یک از سلیقه‌های سیاسی، فرهنگی، و ورزشی یکسان نمی‌باشد. ممکن است کاربر در یک زمان معین، اخبار ورزشی را بر اخبار سیاسی و فرهنگی ترجیح دهد.

با در نظر گرفتن این زوج مولفه، رفتار کاربر با استفاده از یک بردار توزیع احتمالاتی مدل می‌شود. این بردار توزیع شامل زوج‌های <**سلیقه ، تمایل**> می‌باشد که پارامتر "**تمایل**" احتمال انتخاب پارامتر "**سلیقه**" نظیر آن را (به عبارتی اهمیت آن "**سلیقه**" ) در یک زمان معین نشان می‌دهد. با توجه به اینکه مولفه تمایل کاربر در هر زمان محاسبه می‌شود، بدین ترتیب امکان پویایی سلیقه کاربر را فراهم می‌سازد. به عبارتی اولویت سلایق کاربر در طول زمان متغیر می‌باشد. در روش پیشنهادی، از پروفایل هر کاربر (که شامل اطلاعات محتوایی از منابع نیز می‌باشد) برای ایجاد مدل کاربر استفاده می‌شود. بدین ترتیب، تعداد سلیقه به ازای هر کاربر مستقل از دیگران می‌باشد.

[1] - Preference

بنابراین در این رساله ابتدا بررسی و تحلیل رفتار کاربران جهت استخراج زوج مولفه <**سلیقه ، تمایل**> به عنوان عوامل موثر در رفتار کاربر (در انتخاب منابع) انجام می‌شود. سپس از آن در مدلسازی رفتار کاربر و بهبود عملکرد سیستم توصیه‌گر استفاده می‌شود.

## ۱-۴- نوآوری و دستاوردها

در این رساله با استفاده از مولفه‌های سلیقه و تمایل کاربر، یک مدل رفتاری از پروفایل کاربر ایجاد می‌شود. سپس این مدل رفتاری برای ایجاد یک سیستم توصیه‌گر مورد استفاده قرار می‌گیرد. نوآوری اصلی این رساله معرفی مولفه تمایل کاربر و بکارگیری آن در جهت ایجاد مدل مناسب از کاربر می‌باشد. این مولفه می‌تواند پویایی سلیقه کاربر در طول زمان را رصد نماید. برای ایجاد مدل رفتاری کاربر، از چارچوب شبکه بیزین ناپارامتریک (Bayesian non-Parametric framework) به نام مدل ترکیبی فرآیند دریکله (Dirichlet Process Mixture Model) استفاده شده‌است [۳۹-۴۱]. سلیقه و تمایل کاربر به عنوان پارامترهای پنهان در این مدل تعریف می‌شوند. پروفایل کاربر در مدل ترکیبی فرآیند دریکله بکارگرفته می‌شود و با استنتاج پارامترها، سلیقه و تمایل کاربر استخراج می‌شوند. این مدل ترکیبی دارای دو مولفه اساسی می‌باشد. مولفه اول فرآیند رستوران چینی ( Chinese Restaurant Process) می‌باشد و یک خوشه‌بندی افزایشی را ایجاد می‌کند. این خوشه‌بند افزایشی جهت استخراج سلیقه کاربر استفاده می‌گردد. مولفه دوم نیز فرآیند دریکله (Dirichlet Process) می‌باشد و توزیع احتمال خوشه‌ها (سلیقه‌ها) را در قالب تمایل کاربر محاسبه می‌کند. نوآوری این رساله را می‌توان بصورت خلاصه در موارد ذیل بیان نمود:

- در روش پیشنهادی علاوه بر مولفه سلیقه، مولفه **تمایل** برای مدلسازی رفتار کاربر در انتخاب منابع تعریف شده‌است. بدین ترتیب از زوج مولفه <**سلیقه ، تمایل**> برای ایجاد مدل کاربر استفاده می‌شود. مولفه تمایل، میزان اهمیت هر یک از سلایق کاربر در یک زمان معین را نشان

می‌دهد. مولفه سلیقه، امکان بازیابی مجموعه‌ای از منابع بازیابی که احتمالا توسط کاربر انتخاب می‌شوند را فراهم می‌کند. مولفه تمایل این مجموعه منابع را رتبه‌بندی و به کاربر ارائه می‌کند.

- مدل پیشنهادی پویا می باشد و با تغییرات رفتار کاربر در طی زمان در انتخاب منابع سازگاری دارد. در روش پیشنهادی، تعداد سلیقه‌های کاربر در طول زمان متغیر می‌باشند و همراه با تغییرات رفتار کاربر در انتخاب منابع به روز می‌شوند. همچنین مولفه تمایل نیز براساس انتخاب-های کاربر در طول زمان، اهمیت هر سلیقه را به روز می‌کند.
- مدل پیشنهادی برخلاف روش‌های موجود نظیر مدل‌های مبتنی بر فاکتور پنهان که از مولفه سلیقه استفاده می‌کنند، قابلیت پیشنهاد منابعی که اولین بار وارد سیستم می‌شوند (شروع سرد) را دارا می‌باشد ( مشکل منبع جدید ندارد). در روش پیشنهادی، از محتوای منابع در فرآیند مدلسازی رفتار کاربر و استخراج سلایق استفاده می‌شود. بنابراین، با بکارگیری ویژگی محتوایی منابع جدید، بررسی میزان شباهت آنها به سلایق کاربر و امتیازدهی آنها فراهم می‌باشد.
- با توجه به بخش ۱-۲-۱، دقت توصیه‌ها از جمله مسائلی می‌باشد که در روش پیشنهادی به آنها پرداخته شده است. در بخش ارزیابی، توانایی روش پیشنهادی در بهبود عملکرد سیستم‌های توصیه‌گر تایید شده است. نتایج ارزیابی نشان می‌دهد که مولفه تمایل توانسته در تخمین میزان محبوبیت منابع کاندید و رتبه‌بندی آنها موثر باشد.
- مدل پیشنهادی دارای یک ابرپارامتر عددی می‌باشد که توزیع سلیقه‌ها را در طول زمان کنترل می‌کند. این ابرپارامتر نیاز به مقداردهی اولیه ندارد و بصورت خودکار از روی داده‌های ورودی (پروفایل کاربر) محاسبه می‌شود. در روش پیشنهادی، این پارامتر از میانگین شباهت بین منابع انتخاب شده توسط کاربر محاسبه می‌گردد.
- در این رساله، با توجه به مدلسازی رفتار کاربر، یک معیار شباهت برای بهبود عملکرد سیستم‌های توصیه‌گر مبتنی بر پالایش مشارکتی ارائه شده است. روش‌های موجود غالبا از داده‌های خام (نظیر امتیاز کاربران به منابع) برای محاسبه شباهت استفاده می‌کنند که با مشکلات عمده‌ای

مواجه می‌باشند. در روش پیشنهادی با استخراج زوج مولفه <**سلیقه ، تمایل**>، ویژگی‌های سطح بالایی را برای کاربران ایجاد می‌کند. سپس، از این ویژگی‌ها برای محاسبه شباهت بین کاربران و ارائه توصیه بهره می‌برد.

جهت تحقق فرضیه تعریف شده، انجام رساله به سه مرحله تقسیم شده‌است. مرحله اول، شامل تحلیل پروفایل کاربران جهت استخراج زوج مولفه‌های <**سلیقه ، تمایل**> کاربر جهت ایجاد مدل رفتاری می‌باشد. با تحقیقات انجام شده در این مرحله، درستی فرضیه اثبات شده است و نتایج آن در مقاله ذیل آمده‌است.

- Bagher, Rahimpour Cami, Hamid Hassanpour, and Hoda Mashayekhi. "User trends modeling for a content-based recommender system." *Expert Systems with Applications 87 (2017): 209-219.*

در مرحله دوم، اعتبارسنجی و بهبود مدل پیشنهادی انجام شده‌است. مدل پیشنهادی برای ایجاد یک سیستم توصیه‌گر فیلم بکارگرفته شده‌است و نتایج در مقاله ذیل به چاپ رسیده‌است.

- Cami, Bagher Rahimpour, Hamid Hassanpour, and Hoda Mashayekhi. "A content-based movie recommender system based on temporal user preferences." *In Intelligent Systems and Signal Processing (ICSPIS), 2017 3rd Iranian Conference on, pp. 121-125. IEEE, 2017.*

همچنین بهبود ساختاری و الگوریتمی نیز در مدل پیشنهادی ایجاد شده‌است که نتایج در مقالات ذیل آمده‌است.

- Cami Bagher Rahimpour, Hassanpour Hamid, Mashayekhi Hoda, "Inferring user preference using distance dependent chinese restaurant process and weighted distribution for a content based recommender system", *In Artificial Intelligence and Signal Processing Conference (AISP). IEEE, 2017.*

- Bagher, Rahimpour Cami, Hamid Hassanpour, and Hoda Mashayekhi. "User Preferences Modeling using Dirichlet Process Mixture Model for a Content-Based Recommender System." *Knowledge-Based Systems. Accepted: Aug 26 2018.*

در مرحله سوم، مدل پیشنهادی در یک الگوریتم توصیه‌گر ترکیبی بکارگرفته شد. ابتدا با استفاده از رویکرد پالایش محتوایی، پروفایل کاربران جهت ایجاد مدل رفتاری مبتنی‌بر تمایلات کاربران تحلیل شده‌است. سپس با استفاده از معیار شباهت مبتنی‌بر تمایل کاربر (کاربران)، همسایه کاربر متقاضی شناسایی و به کمک رویکرد پالایش مشارکتی، پیشنهاد منابع انجام می‌گیرد. نتایج این مرحله در مقاله ذیل گردآوری شده‌است.

- Bagher, Rahimpour Cami, Hamid Hassanpour, and Hoda Mashayekhi. "A Preference-based User Similarity to Construct a Collaborative Recommender System." *Information Sciences (submitted).*

### ۱-۵- مفروضات

در انجام این رساله، از رویکرد ترکیبی با مولفه‌های پالایش مبتنی‌بر محتوا و پالایش مشارکتی برای ایجاد سیستم توصیه‌گر استفاده می‌شود. بدین ترتیب، باید ویژگی‌های محتوایی منابع موجود باشد. به همین ترتیب از ابزارهای پردازش متن موجود برای ایجاد بردار ویژگی (بردار کلمات) استفاده می‌شود. همچنین هدف رساله ایجاد مدل مناسب کاربر برای بهبود سیستم توصیه‌گر می‌باشد و بهبود الگوریتم ارائه توصیه مدنظر نمی‌باشد.

## ۱-۶- ارزیابی

**داده‌های ارزیابی.** برای ارزیابی مدل پیشنهادی، از دو مجموعه داده استفاده شده‌است. مجموعه اول، داده MovieLens[1] می‌باشد. این مجموعه داده شامل امتیاز کاربران به فیلم‌های انتخاب شده در طول زمان می‌باشد. همچنین شرح فیلم‌ها بصورت داده متنی برای هر فیلم موجود می‌باشد. مجموعه دوم با نام NewsTweet[2]، شامل توئیت‌های خبری است که کاربران مختلف از کانال‌های خبری بازدید نموده‌اند. در این مجموعه داده، به ازای هر کاربر، توئیت‌های خبری بازدید شده از کانال‌های New York Times, BBC, CNN, Reuters, Associated Press گردآوری شده‌است.

**معیارهای ارزیابی.** برای ارزیابی میزان دقت مدل پیشنهادی (میزان مطابقت منابع پیشنهاد شده با انتخاب کاربر) از معیارهای دقت، بازخوانی(precision, recall) و معیار (F1-measure) استفاده شده‌است [۴۲]. همچنین برای تعیین میزان درستی ترتیب لیست منابع پیشنهاد شده از معیار DCG@N[3] استفاده شد [۲۲, ۴۳].

**روش‌های بکارگرفته شده جهت مقایسه.** برای ارزیابی میزان کارائی مدل پیشنهادی و مقایسه آن با روش‌های موجود، تعدادی از روش‌های مشهور انتخاب شده‌اند که عمومیت بیشتری در حوزه سیستم‌های توصیه‌گر دارند و در قالب ابزارها و پیمانه‌های نرم‌افزاری موجود می‌باشند، و عبارتند از LensKit[4]، Surprise[5]، LibRec[6] [۱۸, ۴۴, ۴۵]. در این ابزارها الگوریتم‌های پالایش مشارکتی [۴۶]، پالایش محتوایی [۷]، و تکنیک‌های تجزیه ماتریس [۴۷] پیاده‌سازی شده‌اند.

[1] - https://grouplens.org/datasets/movielens
[2] - http://ipdm.shahroodut.ac.ir/downloads/link
[3] - Discounted Cumulative Gain (DCG)
[4] - http://lenskit.grouplens.org/
[5] - http://surprise.readthedocs.io/en/stable
[6] - https://www.librec.net/

## ۱-۷- ساختار رساله

در ادامه رساله، ابتدا تشریح و تحلیل روش‌های موجود برای توسعه سیستم‌های توصیه‌گر در فصل دوم آمده‌است. در فصل سوم، جزئیات مدل پیشنهادی برای مدلسازی رفتار کاربران در حوزه سیستم-های توصیه‌گر و نحوه بکارگیری آن در ایجاد یک سیستم توصیه‌گر ترکیبی بیان می‌شود. ارزیابی مدل پیشنهادی، بیان نتایج آزمایشات، و تحلیل میزان کارائی روش پیشنهادی در فصل چهارم آمده‌است. فصل پنجم نیز شامل جمع‌بندی و پیشنهاد کارهای آتی می‌باشد.

# فصل دوم

# بررسی روش‌های موجود برای توسعه سیستم‌های توصیه‌گر

## ۲-۱- مقدمه

در تحقیقات انجام شده، براساس پروفایل کاربران (داده‌های ورودی) و نحوه بکارگیری آنها در مکانیسم ارائه توصیه، روش‌های متعددی توسعه یافته‌اند. روش‌های موجود از دو رویکرد **الگوریتمی** و **مدلسازی** پروفایل کاربران، در توسعه سیستم‌های توصیه‌گر استفاده می‌کنند. در رویکرد الگوریتمی انتخاب تکنیک مناسب و مکانیسم ارائه توصیه مدنظر می‌باشد. در رویکرد مدلسازی نیز نحوه مدلسازی و ارائه پروفایل کاربران جهت استفاده در الگوریتم‌های ارائه توصیه انجام می‌شود. در ادامه، تشریح روش‌های موجود آمده‌است.

## ۲-۲- مروری بر رویکرد الگوریتمی در توسعه سیستم‌های توصیه‌گر

با توجه به حوزه کاربرد یا به عبارتی نوع ورودی(ها)، الگوریتم‌های مختلفی برای ایجاد سیستم‌های توصیه‌گر ارائه شده‌اند. در این الگوریتم‌ها به نحوه اعمال تکنیک‌ها بر داده (های) ورودی جهت استخراج سلیقه کاربر و استفاده از آن برای پیشنهاد منابع پرداخته می‌شود. با توجه به شکل۱-۱، حوزه کاربرد، نوع داده(های) ورودی به مولفه محاسباتی را تعیین می‌کند و داده(های) ورودی نیز الگوریتم مناسب برای انجام پردازش لازم در مولفه محاسباتی را مشخص می‌کند. از اینرو بنا به کاربردهای مختلف می‌توان الگوریتم مناسب را انتخاب نمود. جدول ۲-۱ دسته‌بندی کلی روش‌های ارائه شده براساس تکنیک‌ها و داده‌های بکارگرفته شده در هریک از آنها را نشان می‌دهد. در این جدول اولین ستون از سمت راست روش‌های مختلف ارائه شده را نشان می‌دهد. ستون دوم تکنیک-های استفاده شده در هریک از این روش‌ها را بیان می‌کند. ستون سوم نیز، داده(های) ورودی به سیستم توصیه‌گر را نشان می‌دهد. به عنوان مثال، روش پالایش مشارکتی مبتنی بر حافظه با اعمال تکنیک‌های نزدیکترین همسایگی ( به کمک معیارهای شباهت نظیر Pearson و Cosine) بر روی بازخورد کاربران، محاسبات لازم برای ارائه توصیه را انجام می‌دهد. در ادامه جزئیات مربوط به هریک از روش‌های بیان شده در جدول ۲-۱، بررسی می‌شوند.

## الف) سیستم‌های توصیه‌گر مبتنی بر پالایش مشارکتی

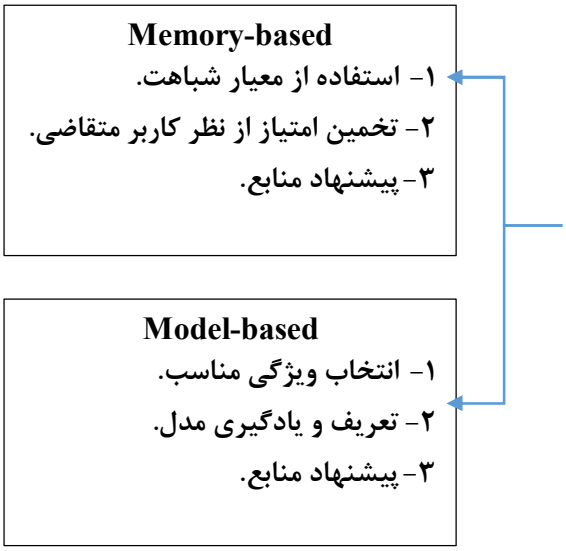
ایده اصلی پالایش مشارکتی این است که، کاربران دارای سلایق مشترک[1]، منابع مشابهی را نیز استفاده می‌کنند. بدین ترتیب، در این روش منابعی به کاربر متقاضی پیشنهاد می‌شوند که کاربران مشابه، استفاده نموده‌اند.

جدول ۲-۱: دسته‌بندی الگوریتم‌های موجود برای سیستم توصیه‌گر براساس تکنیک‌ها و داده ورودی.

| روش‌های ارائه شده | تکنیک‌های بکارگرفته شده جهت پردازش | داده(های) ورودی |
|---|---|---|
| الگوریتم‌های مشارکتی مبتنی بر حافظه (Memory-Based CF) | استفاده از معیارهای شباهت مانند Cosine, Pearson | ماتریس بازخورد کاربران (User's Rating) |
| الگوریتم‌های مشارکتی مبتنی بر مدل (Model-Based CF) | استفاده از تکنیک‌های خوشه‌بندی مانند Bayesian networks, k-means | ماتریس بازخورد کاربران (User's Rating) |
| الگوریتم‌های مبتنی بر محتوا (Content-Based) | تکنیک‌های پردازش متن مانند TF-IDF ( term frequency–inverse document frequency) | ویژگی منابع استفاده شده ( Item's Content) |
| الگوریتم‌های مبتنی بر زمینه تعامل (Context-Based) | استفاده از اطلاعات زمینه تعامل مانند زمان و مکان و انجام پردازش‌های چند بعدی (Multi-level Multidimensional Rating Estimation) | داده‌های موجود در زمینه تعامل (User's Behavior+ Social environment) |
| الگوریتم‌های مبتنی بر پایگاه دانش (Knowledge-Based) | پردازش مبتنی بر قوانین و یا مبتنی بر نمونه (Instance-Based, Rule-Based) | قوانین استفاده از منابع و تعریف نیاز کاربر (Constraints +Critique+ User's Needs) |
| الگوریتم‌های مبتنی بر شبکه اجتماعی (Social-Based) | استفاده از شبکه اجتماعی در پالایش مشارکتی (Social Information + Basic CF) | روابط بین کاربران در شبکه اجتماعی (Social Familiarity+ Trust) |
| الگوریتم‌های مبتنی بر مشخصات کاربر (Demographic) | استفاده از تکنیک‌های خوشه‌بندی | ویژگی افراد (Demographic Data) |

[1] - Like Minded

این روش بر پایه امتیازدهی انجام شده توسط کاربران برای هر منبع عمل می‌نماید [۲۶, ۲۸]. در این روش، امتیاز کاربر به ازای هر منبعی که استفاده نموده‌است، در ماتریس بازخورد[1] ذخیره می‌شود. هر درایه از این ماتریس بیان کننده امتیاز یک کاربر خاص(سطر ماتریس) به یک منبع خاص(ستون ماتریس) می‌باشد. در شکل ۲-۱، نمونه‌ای از این ماتریس آمده‌است. در این روش به کمک ماتریس بازخورد، مجموعه‌ای از کاربران مشابه، یا در اصطلاح، همسایگی[2] یک کاربر شناسایی می‌شوند. سپس براساس منابعی که این کاربران مشابه استفاده نموده‌اند، پیشنهاد منابع جدید انجام می‌شود. این ماتریس بازخورد به روش‌های مختلفی مورد استفاده قرار گرفته‌است. این روش‌ها که مبتنی بر ماتریس بازخورد می‌باشند را می‌توان با توجه به شکل ۲-۱ به دو دسته تقسیم نمود [۲۷, ۲۸]. دسته اول به روشهای مبتنی بر حافظه (مبتنی بر شباهت)[3] و دسته دوم به روشهای مبتنی بر یادگیری[4] معروف می‌باشند. در ادامه ضمن شرح دو دیدگاه به همراه روش‌های ارائه شده در هر یک از آنها بیان می‌شوند.




| | فیلم ۱ | فیلم ۲ | فیلم ۳ | فیلم ۴ | فیلم ۵ |
|---|---|---|---|---|---|
| کاربر ۱ | ۵ | ۳ | ۴ | ۴ | ؟ |
| کاربر ۲ | ۳ | ۳ | ۲ | ۳ | ۲ |
| کاربر ۳ | ۴ | ۳ | ۴ | ۳ | ۵ |
| کاربر ۴ | ۳ | ۳ | ۱ | ۵ | ۴ |
| کاربر ۵ | ۱ | ۵ | ۵ | ۲ | ۱ |

شکل ۲-۱: شمای کلی سیستم‌های توصیه‌گر مبتنی بر فیلترینگ مشارکتی.

[1] - Utility Matrix or Feedback Matrix
[2] - Nearest Neighbor
[3] - Memory-based (Neighborhood-based Methods)
[4] - Model-based

### ❖ پالایش مشارکتی مبتنی بر شباهت

در این دیدگاه با بکارگیری معیارهای شباهت، کاربران (منابع) مشابه شناسایی می‌شوند. سپس براساس کاربران (منابع) مشابه (همسایه)، پیشنهاد منابع به کاربر متقاضی انجام می‌شود. باتوجه به ماتریس بازخورد و معیار شباهت، مراحل ذیل برای توصیه منابع مناسب به کاربر متقاضی انجام می‌شود [۴۶, ۴۸, ۴۹]. غالبا در پالایش مشارکتی مبتنی بر شباهت از معیارهای شباهت Cosine, Pearson, Jaccard, و mean square difference [۵۰, ۵۱] استفاده می‌شود.

در این دیدگاه با توجه به اینکه در انجام محاسبات، هدف یافتن کاربران مشابه کاربر یا منابع مشابه باشد، دو روش ایجاد شده‌است [۹]. در ادامه این دو روش بررسی می‌شوند.

### ❖ پالایش مشارکتی مبتنی بر شباهت کاربران

این روش بر مبنای منابع توصیه شده از طرف کاربران مشابه می‌باشد. ابتدا گروه کاربران مشابه کاربر جاری شناسایی می‌شوند. سپس منابع استفاده شده توسط این گروه به عنوان نتیجه به کاربر جاری توصیه می‌شوند [۴۸].

با توجه به مطالعات انجام شده، مقاله [۵۲] به عنوان نخستین سیستم توصیه‌گر مبتنی بر کاربر به نام Tapestry می‌باشد. این سیستم به صورت پیش فرض برای پالایش متون الکترونیک پیاده‌سازی شده‌است. در این مقاله یک معماری دولایه برای سیستم توصیه‌گر ارائه شده‌است. لایه اول سمت سرور می‌باشد که متون الکترونیک و بازخوردها نیز در این لایه ذخیره می‌شوند. در این لایه در مرحله اول با انجام پردازش‌های اولیه روی منابع، بردار ویژگی از منابع تولید و ذخیره می‌گردد. سپس در مرحله دوم مکانیسم پالایش منابع و گردآوری منابع مناسب کاربر انجام می‌شود. در لایه دوم که سمت کاربر می‌باشد، یک ماژول اولویت‌بندی جهت نمایش منابع و یک واسط کاربر جهت دریافت بازخورد کاربر و پارامترهای پرس‌وجو قرار دارند. به کمک واسط کاربر، درخواست کاربر به صورت یک پرس‌وجو به ماژول پالاینده در سمت سرور ارسال می‌شود. در ساخت این پرس‌وجو، از بازخورد

کاربران دیگر استفاده می‌شود. به عنوان مثال "منابعی که کاربر X استفاده کرده و امتیاز ۱۰ را به آن نسبت داده‌است." یک پرس‌وجویی می‌باشد که از سمت کاربر به ماژول پالاینده ارسال می‌گردد. در حقیقت ایده اصلی این مقاله در ساخت این پرس‌وجو می‌باشد. ماژول پالاینده با اعمال پرس‌وجو، منابع مناسب را به کمک بردار ویژگی منابع و بازخوردها استخراج می‌کند.

هر چند در مقاله [۵۲] از بازخورد کاربران برای پالایش استفاده شده‌است اما رابطه بین کاربر و منبع هدف، به صورت صریح مدل نشده‌است. توسعه‌های نیز بر این کار انجام شده‌است. از جمله ویژگی بارز این توسعه‌ها بیان صریح رابطه بین کاربر و منبع می‌باشد که به صورت یک ماتریس بازخورد مدل شده‌است. در این توسعه‌ها، برای ارائه توصیه به کاربر متقاضی، ابتدا به کمک معیارهای شباهت، میزان تشابه کاربر متقاضی با سایر کاربران ارزیابی می‌شود. سپس با استفاده از میزان شباهت هر کاربر (n) به کاربر متقاضی (u)، میزان علاقه کاربر متقاضی به یک منبع خاص (i) پیش‌بینی می‌شود. سپس لیستی از منابع مرتب شده براساس مقادیر پیش‌بینی شده، ارائه می‌شود [۹, ۴۵, ۵۳, ۵۴].

رابطه ۲-۱و ۲-۲، به ترتیب نحوه محاسبه شباهت کاربران ($Sim(u,n)$) و پیش‌بینی میزان علاقه کاربر متقاضی به یک منبع خاص ($Pred(u,i)$) را نشان می‌دهند. در رابطه ۳-۱، از معیار شباهت Pearson استفاده شده‌است [۵۳].

$$Sim(u,n) = \frac{\sum_{i\in R_{u,n}}(r_{ui}-\overline{r_u})(r_{ni}-\overline{r_n})}{\sqrt{\sum_{i\in R_{u,n}}(r_{ui}-\overline{r_u})^2}\sqrt{\sum_{i\in R_{u,n}}(r_{ni}-\overline{r_n})^2}} \quad (۱-۲)$$

$$Pred(u,i) = \frac{\sum_{n\in \text{neighbors(u)}} Sim(u,n) * r_{ni}}{\sum_{n\in \text{neighbors(u)}} Sim(u,n)} \quad (۲-۲)$$

از آنجاییکه در این روش تمام پردازش‌ها به صورت همزمان با درخواست کاربر می‌باشد از این رو پیچیدگی محاسباتی بالایی دارد. به عنوان نمونه تجاری، در مقاله [۴۵] از الگوریتم پالایش مشارکتی برای ایجاد یک سیستم توصیه‌گر خبر استفاده شده‌است. کاربران به ازای استفاده از هر خبر، یک

امتیاز به آن نسبت می‌دهد. با استفاده از بازخورد کاربران به اخبار مطالعه شده، ماتریس بازخورد ایجاد می‌شود. سپس این ماتریس بازخورد با استفاده از الگوریتم پالایش مشارکتی پردازش می‌شود و براساس آن توصیه خبر انجام می‌شود.

### ❖ پالایش مشارکتی مبتنی بر شباهت منابع

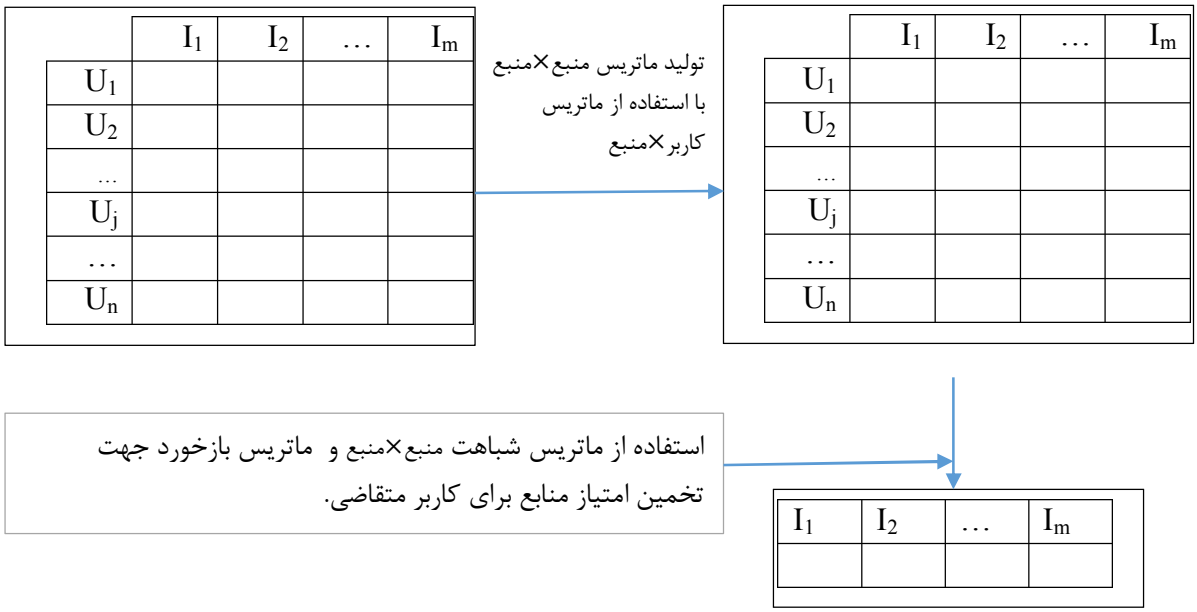
با وجود اینکه روش مبتنی بر کاربر، نتایج مناسبی تولید می‌کنند اما مشکلاتی مانند خلوت بودن ماتریس و پیچیدگی محاسباتی سبب شد تا روش مبتنی بر منبع نیز ارائه گردد. در این روش براساس امتیازی که کاربر به منابع استفاده شده می‌دهد، منابع مشابه را شناسایی و امتیازدهی می‌کند.

مولفان در مقاله [۴۶] یک روش‌های مبتنی بر منبع ارائه نمودند که برای تعیین منابع مشابه، ماتریس بازخورد M ایجاد می‌شود. این ماتریس به ابعاد تعداد منابع می‌باشد و Mij بیان کننده میزان شباهت دو منبع i و j می‌باشد. برای میزان شباهت نیز یک مقدار حد آستانه تعیین می‌شود و مقادیر کمتر از حد آستانه، صفر می‌شوند. برای تعیین شباهت از دو معیار Cosine و احتمال شرطی(بیز) استفاده شده‌است. با ایجاد ماتریس M، براساس منابعی که کاربر متقاضی استفاده کرده، منابع مشابه در M را یافته و به کاربر پیشنهاد می‌کند. در این مقاله توسعه‌هایی نیز بر روش مبتنی بر منبع داده شده که سیستم را قادر می‌سازد در توصیه‌ها به جای بررسی احتمال استفاده توام دو منبع، احتمال استفاده توام یک مجموعه از منابع به کاربر پیشنهاد شود.

در مقاله [۵۵] نیز ابتدا به کمک ماتریس بازخورد، شباهت منابع محاسبه شده‌است. سپس به کمک الگوریتم مجموع وزن‌دار، پیشنهاد منبع مطابق رابطه ۲-۳، انجام می‌شود. در این رابطه $P_{u,i}$، پیش‌بینی کاربر u از منبع استفاده نشده i و $R_{u,N}$، بازخورد کاربر u نسبت به منبع استفاده شده N می‌باشند. در این مقاله ضمن توجه به مشکلاتی که معیارهای شباهت دارند، در مرحله پیش بینی به جای استفاده مستقیم از امتیاز منابع مشابه، از امتیاز تعدیل شده منابع مشابه به کمک الگوریتم رگرسیون برای بکارگیری در الگوریتم مجموع وزن‌دار استفاده می‌شود. شکل ۲-۲، مراحل انجام روش مبتنی بر منبع را نشان می‌دهد. با توجه به این شکل ابتدا، به کمک ماتریس بازخورد و معیارهای

شباهت، ماتریس M (که شباهت منابع را نشان می‌دهد) ایجاد می‌شود. سپس به ازای درخواست کاربر متقاضی، منابع مشابه منابع استفاده شده توسط آن را بازیابی و نمایش می‌دهد. ثابت بودن نسبی شباهت منابع از جمله مواردی است که در این روش بر آن تاکید می‌شود و این روش را قادر می‌سازد تا قسمت تعیین شباهت را به صورت غیرهمزمان انجام دهد.

$$P_{u,i} = \frac{\sum_{N \in Similar\ Items} M_{i,N} * R_{u,N}}{\sum_{N \in Similar\ Items} |M_{i,N}|} \qquad (۳-۲)$$




شکل ۲-۲: مراحل پالایش مشارکتی مبتنی بر منبع را نشان می‌دهد.

این الگوریتم برای پایگاه داده‌های بزرگ نظیر Amazon.com نیز استفاده شده‌است [۱]. با توجه به پیاده‌سازی انجام شده در این مقاله، از آنجاییکه این الگوریتم یک مرحله غیرهمزمان برای ساخت ماتریس شباهت دارد از این رو هزینه محاسباتی آن نسبت به روش مبتنی بر کاربر بهتر می‌باشد.

### ❖ پالایش مشارکتی مبتنی بر مدل‌سازی

در این دیدگاه ابتدا براساس ماتریس بازخورد، رابطه بین کاربر و منابع به کمک الگوریتم‌های یادگیری مدل می‌شود. سپس براساس مدل ایجاد شده، پیش بینی صورت می‌گیرد. در این دیدگاه، مسئله پالایش مشارکتی به کمک الگوریتم‌هایی نظیر شبکه‌های بیزین، خوشه‌بندی، مدل‌های

احتمالاتی انجام می‌شود [۴۶, ۵۵]. از جمله نکته اساسی که در این دیدگاه تاکید شده‌است، غلبه بر مشکل پیچیدگی محاسباتی می‌باشد زیرا این دسته از روش‌ها دارای یک فاز غیرهمزمان می‌باشند [۴۶]. مقاومت در برابر حمله‌ها نیز از نتایج استفاده از این دسته از روش‌ها می‌باشد [۵۶]. در ادامه برخی از روش‌های معمول مدلسازی استفاده شده در این دیدگاه بیان می‌شوند.

### ❖ پالایش مشارکتی مبتنی بر یادگیری به کمک تئوری بیز

در این روش به کمک تئوری بیز، مدلی از ماتریس بازخورد استخراج می‌شود. سپس به کمک مدل ایجاد شده، منابع مناسب را به کاربر متقاضی پیشنهاد می‌گردد $58٪. در این روش امتیاز یک منبع $I_i$ به صورت یک احتمال شرطی مطابق با تئوری بیز مدل می‌شود. در این روش جهت یافتن منابع مناسب برای کاربر $U_j$، امتیاز هر منبع $I_i$ از نظر کاربر $U_j$ بصورت احتمال شرطی $P(I_i|U_j)$ بیان می‌شود.

با فرض اینکه $R_x^j \in \{0,1\}$ بیان کننده انتخاب یا عدم انتخاب منبع $I_x$ توسط کاربر $U_j$ باشد و $Y = \{R_1^j, \dots, R_{Lj}^j\}$ (Lj، تعداد منابع کاربر را نشان می‌دهد) بیان کننده نتیجه بازخورد منابع استفاده شده باشد، به این ترتیب احتمال انتخاب منبع $I_x$ توسط کاربر $U_j$ با فرض موجود بودن Y، به صورت احتمال شرطی $P(R_x^j \mid Y)$ بیان می‌شود [۵۷].

در حقیقت $P(I_i|U_j)$ را می‌توان معادل همان گروه‌های کاربری مشابه در پالایش مشارکتی در نظر گرفت. در این احتمال کاربرانی که یک منبع مشابه را انتخاب نموده‌اند مشارکت دارند. رابطه ۲-۴، فرم کلی استفاده شده از تئوری بیز در سیستم‌های توصیه‌گر را نشان می‌دهد [۵۸-۶۰]. در این رابطه مقادیر $P(I_i)$ و $P(U_j|I_i)$ به کمک داده‌های آموزشی محاسبه می‌شوند. البته بحث استقلال شرطی مطرح شده در بیز ساده، با ایده اصلی سیستم‌های توصیه‌گر که کاربران مشابه، منابع مشابه استفاده می‌کنند تناقض دارد.

$$P(I_i|\, U_{1..n}) = P(I_i) * \prod_{j:1..n} P(U_j\,|I_i) = P(R_x^j\,) * \prod P(R_{lj}^j|R_x^j) \qquad (۴-۲)$$

از تئوری (یادگیر) بیز برای تخمین درایه‌های خالی ماتریس بازخورد جهت برطرف کردن مشکل تنک بودن ماتریس بازخورد نیز استفاده می‌شود [۶۱].

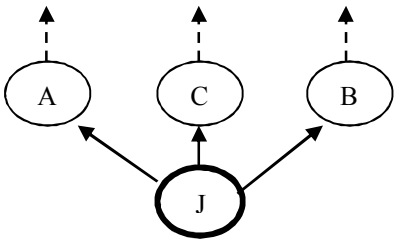
در مقاله [۶۲] الگوریتمی برای ترکیب بیز و پالایش مشارکتی ارائه شده‌است. در این مقاله به کمک الگوریتم ارائه شده، در هر ارزیابی، مقادیر تخمینی توسط دو الگوریتم بیز و پالایش مشارکتی محاسبه می‌شوند. سپس هر یک که نتیجه بهتری ارائه نمایند به عنوان معیار ارزیابی نهایی لحاظ می-شوند. البته برای کاهش بار محاسباتی، پیشنهاد شده‌است که عملیات مربوط الگوریتم بیز به صورت غیرهمزمان انجام شود.

### ❖ پالایش مشارکتی مبتنی بر یادگیری به کمک شبکه بیزی

با توجه به فرض اساسی سیستم‌های توصیه‌گر مبتنی بر پالایش مشارکتی، که کاربران مشابه، احتمالا منابع مشابه را انتخاب می‌کنند، از اینرو می‌توان از احتمال شرطی برای مدل کردن رفتار کاربران استفاده نمود. از آنجاییکه ارتباط بین کاربران یک گراف را ایجاد می‌کند، بنابراین از شبکه‌های بیز به عنوان یک ابزار مناسب برای مدل کردن ارتباط بین کاربران استفاده شده‌است. این شبکه علاوه بر اینکه ارتباط بین سلایق کاربران را مدل می‌کند، می‌تواند شباهت سلیقه کاربران به صورت احتمال شرطی در لینک‌ها نشان دهد. ارتباط بین منابع و ارزیابی کاربران و نیز شباهت کاربران، به صورت لینک‌ها و مقادیر احتمال بیان می‌شوند. در ادامه به چند نمونه از کارهای ارائه شده در این زمینه اشاره می‌شود.

شکل ۲-۳، قسمتی از ارتباط بین کاربران در شبکه اجتماعی را نشان می‌دهد که به کمک شبکه بیز مدل شده‌است. با توجه به شکل ۲-۳، اگر کاربر J، منبع m را درخواست نماید در اینصورت درخواست به کاربران A,B,C که با J در ارتباط هستند ارسال می‌شود. نتایج برگشتی که ارزیابی این کاربران نسبت به منبع m می‌باشد، جمع‌آوری می‌گردد. برای ترکیب این نتایج و بدست آوردن یک مقدار ارزیابی برای منبع m، از استنتاج احتمال در شبکه بیز استفاده می‌گردد. در رابطه ۲-۵، نحوه

استنتاج احتمال شرطی از شبکه بیز (شکل۲-۳) آمده‌است. مقادیر این احتمال‌های شرطی به کمک داده‌های آموزشی محاسبه می‌شود [۵۸].




شکل ۲-۳: قسمتی از ارتباط بین کاربران در شبکه اجتماعی
که به کمک شبکه بیز مدل شده‌است را نشان می‌دهد.

$$P(R_j = r_J | r_A, r_B, r_C) \xrightarrow[\{(R_A \perp R_B \perp R_C \mid R_J\}]{\text{Conditionally Independent}} = P(r_A, r_B, r_C \mid r_J) * P(r_J) \quad (۵-۲)$$

در مقاله [۶۳] ارائه سیستم توصیه‌گر با استفاده از یک شبکه بیز جهت توصیه فیلم بیان شده است. در این مقاله، ابتدا ویژگی‌های کاربر (u)، فیلم (c)، و زمینه تعامل(s) ایجاد می‌شوند. سپس مقدار امتیاز به کمک احتمال شرطی P(v| u,c,s) تخمین زده می‌شود. با توجه به اینکه حوزه کاری در خصوص توصیه فیلم می‌باشد، ساختار شبکه بیز به صورت دستی ساخته می‌شود و سپس مقادیر احتمال براساس داده‌های آموزشی محاسبه می‌شوند. در ایجاد ساختار شبکه بیز در برخی روش‌ها ساختار شبکه بیز توسط خبره تعیین و مقادیر احتمال به روش بیشینه‌سازی امید ریاضی[1] محاسبه می‌شوند [۶۴]. در برخی کارها نیز، شبکه بیز به صورت ساختار سه لایه‌ای ارائه شده است. لایه اول شامل محتوای منبع می‌باشد. لایه دوم شامل کاربرانی است که به محتوای منبع امتیاز می‌دهند. لایه سوم شامل گره‌هایی می‌باشد که کاربران با سلیقه مشترک را نشان می‌دهد [۶۵].

### ❖ پالایش مشارکتی مبتنی بر یادگیری به کمک تئوری احتمالاتی

در این روش، از قوانین احتمال برای استنتاج سلیقه کاربر و پیشنهاد منبع استفاده می‌شود. در این روش نحوه استنتاج مدل شده است و هدف یادگیری داده‌ها و مدل نمودن سلیقه کاربر نمی‌باشد.

[1] - Expectation Maximization

در روش ارائه شده در مقاله [۶۶] ضمن تاکید بر دو مشکل "حجم محاسبات" و "کاربر جدید"، همانند روش‌های مبتنی بر حافظه، بنا به درخواست (آنلاین) محاسبه احتمال جهت استنتاج صورت می‌گیرد. برای بهبود زمان محاسبات به جای بررسی کل ماتریس بازخورد، زیرمجموعه‌ای از این ماتریس استفاده می‌شود. برای غلبه بر مشکل "کاربر جدید"، ابتدا یک زیر مجموعه کوچکی از منابع (حدود ۱۰ منبع) به کاربر پیشنهاد می‌شود تا سلیقه کاربر تخمین زده شود و سپس براساس سلیقه تخمینی، منابع پیشنهاد می‌شوند. البته تعیین این مجموعه کوچک نیز با استفاده از روابط خاصی انجام می‌گیرد که بتواند بهترین تخمین را داشته باشد.

در مراجع [۶۷, ۶۸] نیز، مراحل استنتاج به کمک قوانین احتمال(PLSA[1]) بیان شده است و از تکنیک EM[2] برای محاسبه پارامترها استفاده شده است.

### ❖ پالایش مشارکتی مبتنی بر یادگیری به کمک خوشه‌بندی

با توجه به هزینه محاسباتی بالا در الگوریتم‌های پالایش مشارکتی، یکی از روش‌های کاهش هزینه محاسباتی، استفاده از الگوریتم‌های دسته‌بندی[3] می‌باشد [۱, ۹, ۶۹]. این روش شامل دو فاز دسته-بندی(یادگیری) و بکارگیری می‌باشد. در فاز یادگیری با اعمال الگوریتم‌های دسته‌بندی روی ماتریس بازخورد ابتدا کاربران به دسته‌هایی مجزا تقسیم می‌شوند. سپس در فاز بکارگیری، ابتدا تعیین می‌شود کاربر متقاضی به کدام دسته متعلق می‌باشد و سپس براساس داده‌های آن دسته(قسمتی از ماتریس بازخورد که متعلق به این دسته می‌باشد)، پیش‌بینی می‌شود. در این روش‌ها فاز اول(دسته‌بندی) به صورت غیرهمزمان (آفلاین) انجام می‌شود از اینرو هزینه محاسبات در انجام پیش‌بینی کم می‌باشد، اما از آنجاییکه بخشی از رکوردهای ماتریس بازخورد استفاده می‌شود، درستی توصیه‌ها نسبت به پالایش مشارکتی ضعیف‌تر می‌باشد.

---

[1] - Probabilistic Latent Semantic Analysis
[2] - Expectation Maximization
[3] - Clustering

### ❖ پالایش مشارکتی مبتنی بر یادگیری به کمک تجزیه ماتریس

در روش‌های پالایش مشارکتی ماتریس بازخورد بدون در نظر گرفتن پارامترهای موثر در رابطه بین کاربر و منبع مورداستفاده قرار می‌گیرند. به عنوان مثال کاربر u1 به منبع Ii امتیاز n1 را نسبت می‌دهد. اما فرض کاربر یا معیار کاربر در امتیازدهی معین نمی‌باشد. در مقاله [۸] روشی مبتنی بر تجزیه ماتریس ( Matrix-Factorization) ارائه شده که سعی دارد این معیارهای پنهانی[1] را مدل نماید. در این روش ماتریس بازخورد $R_{n\,x\,m}$ به دو ماتریس $U_{n\,x\,f}$ و $V_{f\,x\,m}$ تجزیه می‌شود. پارامتر f بیان‌کننده تعداد معیارهای پنهان می‌باشد. ماتریس U شامل ویژگی‌های کاربران به ازای هر معیار است و ماتریس V شامل ویژگی‌های منابع به ازای هر معیار می‌باشد. در حالت ایده‌ال با محاسبه U و V باید رابطه ۲-۶ برقرار باشد.

$$R_{n\,x\,m} = U_{n\,x\,f} \cdot V_{f\,x\,m} \qquad (۶-۲)$$

البته در عمل این تساوی با یک مقدار خطا مواجه می‌باشد. در مقاله [۸] محاسبه ماتریس‌های U و V به کمک الگوریتم‌های یادگیری مانند gradient descent stochastic انجام می‌شود. با داشتن ماتریس‌های U وV ماتریس بازخورد R بازسازی می‌شود. در این بازسازی درایه‌های خالی ماتریس بازخورد – که نشان دهنده عدم وجود ارزیابی کاربر برای یک منبع می‌باشد – پیش‌بینی می‌شوند. به عنوان مثال شکل ۲-۴ نحوه محاسبه ماتریس‌های U و V را نشان می‌دهد [۷۰]. تصویر a از این شکل، ماتریس بازخورد اولیه می‌باشد و تصویر b، ماتریس پیش‌بینی شده براساس مقادیر ماتریس‌های U و V می‌باشد. مقادیر درایه‌هایی از ماتریس تصویر b که پررنگتر می‌باشند، مقادیر پیش‌بینی شده را نشان می‌دهند.

البته تعیین مقدار f که بیان کننده تعداد معیارها می‌باشد و یا به عبارتی نحوه محاسبه U و V که تخمین مناسبی از R را ایجاد نمایند و در برابر خلوت بودن ماتریس ورودی مقاوم باشند از مسائل مهم این روش می‌باشند.

---

[1] - Latent Factor

| | $I_1$ | $I_2$ | $I_3$ | $I_4$ | $I_5$ | $I_6$ | $I_7$ | $I_8$ |
|---|---|---|---|---|---|---|---|---|
| $U_1$ | 5 | 2 | **2.5** | 3 | **4.8** | 4 | **2.2** | **4.8** |
| $U_2$ | 4 | 3 | **2.4** | **2.9** | 5 | **4.1** | **2.6** | **4.7** |
| $U_3$ | 4 | **1.7** | 2 | **3.2** | **3.9** | **3.0** | 2 | 4 |
| $U_4$ | **4.8** | **2.1** | **2.7** | **2.6** | **4.7** | **3.8** | **2.4** | **4.9** |
| $U_5$ | 5 | 1 | 2 | **3.4** | 4 | 3 | **1.5** | **4.6** |
| $U_6$ | 4 | 3 | **2.9** | 2 | 4 | **3.4** | 3 | 5 |

(a) User-item matrix

| | $I_1$ | $I_2$ | $I_3$ | $I_4$ | $I_5$ | $I_6$ | $I_7$ | $I_8$ |
|---|---|---|---|---|---|---|---|---|
| $U_1$ | 5 | 2 | | 3 | | 4 | | |
| $U_2$ | 4 | 3 | | | 5 | | | |
| $U_3$ | 4 | | 2 | | | | 2 | 4 |
| $U_4$ | | | | | | | | |
| $U_5$ | 5 | 1 | 2 | | 4 | 3 | | |
| $U_6$ | 4 | 3 | | 2 | 4 | | 3 | 5 |

(b) predicted user-item

| $I_1$ | $I_2$ | $I_3$ | $I_4$ | $I_5$ | $I_6$ | $I_7$ | $I_8$ |
|---|---|---|---|---|---|---|---|
| 1. | -.05 | -.2 | .26 | 1.2 | .54 | -.3 | .52 |
| .19 | -.89 | -.7 | .05 | .68 | .02 | -.6 | .7 |
| .49 | .89 | -.05 | -.6 | .12 | .08 | .02 | 1.6 |
| -.4 | .7 | .27 | -.2 | .99 | .44 | .39 | .74 |
| 1.4 | -1 | .06 | .06 | .23 | .01 | -.3 | .8 |

V

| $U_1$ | $U_2$ | $U_3$ | $U_4$ | $U_5$ | $U_6$ |
|---|---|---|---|---|---|
| 1.5 | 1.2 | 0.3 | 0.8 | 0.6 | -0.01 |
| 0.3 | 0.9 | 1.2 | 0.3 | 1.1 | 0.2 |
| 0.5 | 0.2 | 0.1 | 0.8 | 0.2 | 1.5 |
| 0.3 | 1.3 | -0.4 | 0.7 | -0.9 | 0.6 |
| 1.0 | 1.0 | 0.1 | 1.1 | 1.8 | 0.4 |

U

شکل۲-۴: نحوه محاسبه ماتریس‌های U و V را نشان می‌دهد. تصویر a، ماتریس بازخورد اولیه و تصویر b، ماتریس پیش‌بینی شده را نشان می‌دهند [۷۰].

از آنجاییکه سیستم‌های توصیه‌گر مبتنی بر یادگیری، فرایند یادگیری را به صورت غیر همزمان (آفلاین) انجام می‌دهند، در هر دوره زمانی مجددا آموزش انجام می‌شود. بنابراین برای محیط‌هایی که تغییرات سلیقه کاربر کمتر باشد مناسب است و در صورتیکه تغییرات سلیقه کاربر زیاد باشد، این روش مناسب نمی‌باشد.

صرفنظر از روش‌های مختلف ارائه شده برای الگوریتم مبتنی بر حافظه، باتوجه به این باور اجتماعی که افراد مشابه از نظر سلیقه معمولا ارزیابی مشابهی در مورد منابع دارند، سعی دارد در انتخاب منبع، سلیقه کاربر را مدنظر قرار دهد. این روش تا حد زیادی توانسته مدل مناسبی از سلیقه کاربر متقاضی ایجاد نماید. این روش به دانش قبلی در مورد جزئیات منابع و کاربران نیاز ندارد که در نتیجه برای هر ساختار منبع می‌توان از این روش استفاده نمود [۲۵, ۴۹]. اما نیاز به داده‌های زیاد جهت ایجاد مدل، ثابت بودن بستر تعامل در همه شرایط و واکنش کند به تغییرات سلیقه کاربر از جمله مشکلات اساسی این روش می‌باشد که سبب افزایش هزینه محاسباتی و کاهش مقیاس‌پذیری و حساسیت به

منبع و کاربر جدید می‌شوند. با توجه به آزمایشات انجام شده، سیستم‌های توصیه‌گر مبتنی بر مدل یادگیری در برابر حملات نیز مقاومت بیشتری دارند [۵۶].

**ب) سیستم‌های توصیه‌گر مبتنی بر محتوا**

در این روش، براساس توصیف منبع و ویژگی کاربر، منابع جدید پیشنهاد می‌شوند [۱۰]. منظور از ویژگی کاربر در این روش، توصیف منابعی می‌باشد که کاربر متقاضی قبلا استفاده نموده‌است. بنابراین در این روش بر تاریخچه سلیقه کاربر متقاضی یا به عبارتی بر سابقه تعاملات متقاضی تاکید می‌شود. در حالیکه در الگوریتم مبتنی بر حافظه[1] بر سابقه تعاملات متقاضی و سلیقه افراد مشابه تاکید می‌شود. به عنوان مثال در یک سایت ارائه خبر، اگر کاربر متقاضی در N مراجعه قبلی از اخیار ورزشی بازدید نمود، بنابراین در درخواست جاری نیز به احتمال خیلی زیاد اخبار ورزشی از انتخاب‌های اولیه این کاربر می‌باشند.

با توجه به گوناگونی در ماهیت منابع، نحوه تعریف و ارائه منابع یکی از مسائل مهم در این روش می‌باشد. معمولا منابع از نظر ماهیت وجودی به دو دسته تقسیم می‌شوند. دسته اول: منابعی که با ساختار معین قابل تعریف می‌باشند. به عنوان مثال، در ارائه یک کتاب می‌توان، ویژگی‌هایی مانند مولف، عنوان و قیمت را برای تعریف محتوایی آن استفاده نمود و یا به عبارتی به صورت یک جدول پایگاه داده مدل نمود. دسته دوم: منابع بدون ساختار هستند. به عنوان مثال خبر، به عنوان یک منبع بدون ساختار متنی می‌باشد که شامل مجموعه‌ای از کلمات است.

صرفنظر از نوع منابعی که استفاده می‌شوند، ویژگی‌های منابع استفاده شده به عنوان ویژگی کاربر ذخیره می‌شوند. به کمک الگوریتم‌های یادگیری مانند نزدیکترین همسایه، درخت تصمیم، دسته‌بند خطی، روش‌های احتمالاتی (بیز)، ویژگی کاربر مدل می‌شود. این مدل همانند یک تابع پیش‌بینی عمل می‌کند. به عبارتی به ازای منبع جدید، میزان احتمالی علاقه کاربر به این منبع را به‌دست می‌دهد [۱۰, ۷۱].

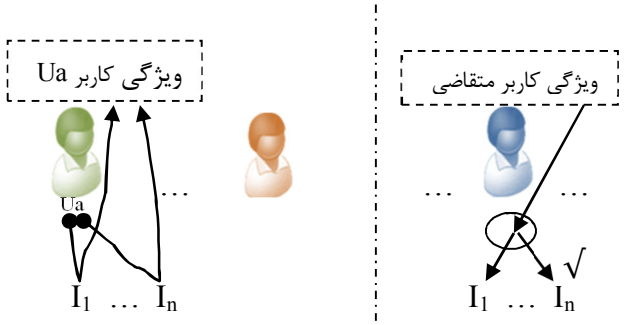
[1] - Memory-based Collaborative Filtering

در مواردی که تعداد منابع استفاده شده کاربر زیاد باشد در اینصورت حجم ویژگی کاربر بسیار زیاد می‌شود و هزینه محاسباتی جهت پردازش ویژگی نیز افزایش می‌یابد [۱]. شکل ۲-۵، نمای کلی این روش را نشان می‌دهد. در قسمت الف از شکل۲-۵، براساس منابع استفاده شده، ویژگی کاربر ساخته می‌شود و تابع پیش‌بینی مدل می‌گردد. در قسمت ب از شکل ۲-۵، براساس ویژگی ایجاد شده، انتخاب منبع جدید انجام می‌شود.

در مقاله [۷۲] از رابطه ۲-۷ برای تعیین میزان شباهت یک منبع به کاربر متقاضی براساس ویژگی کاربر و ویژگی منبع استفاده شده است. در این رابطه تابع $Prof(u_x)$، ویژگی‌های منابعی که کاربر $u_x$ استفاده کرده را بازیابی می‌کند. و $Prof(i_n)$، ویژگی منبع $i_n$ را بدست می‌دهد.

$$Sim(u_x, i_n) = \frac{2 * |Profile(u_x) \cap Profile(i_n)|}{|Profile(u_x)| + |Profile(i_n)|} \qquad (۲-۷)$$

از جمله مزایای این روش، مستقل از کاربر بودن، توصیه در مورد منبع جدید می‌باشد. از جمله معایب این روش، محدودیت در تحلیل محتوایی منبع(استخراج ویژگی) می‌باشد.




الف) ایجاد ویژگی به کمک محتوای منابع استفاده

ب) تعیین منبع متناسب به کمک

شکل۲-۵: فرایند انجام توصیه در الگوریتم مبتنی بر محتوا را نشان می‌دهد.

به عنوان نمونه در مقاله [۷۳]، الگوریتم مبتنی بر محتوا برای توصیه خبر ارائه شده است. در این مقاله ابتدا یک خوشه بندی چند سطحی از اخبار جدید انجام می‌شود. ویژگی کاربر نیز بصورت سه تایی {topic distribution، similar access pattern ، Named Entities} ایجاد می‌شود. که topic distribution بیان‌کننده کلمات کلیدی اخبار مطالعه شده می‌باشد. similar access pattern نیز

بیان‌کننده کاربران مشابه کاربر جاری می‌باشد و Named Entitie بیان‌کننده نام‌ها و قیدهای زمان و مکان موجود در خبر می‌باشد. ویژگی خبر نیز شامل این سه‌تایی می‌باشد. در این مقاله در مرحله اول به کمک ویژگی کاربر که شامل اخبار مطالعه شده می‌باشد، خوشه‌های خبری مشابه ویژگی کاربر متقاضی استخراج می‌شوند. با انتخاب اخبار یک خوشه در مرحله دوم مطلوبیت اخبار این خوشه از نظر کاربر بررسی می‌شود. در این روش از کل ویژگی کاربر(topic distribution) که در طول زمان ایجاد شده است برای تعیین شباهت استفاده می‌گردد. به این ترتیب تغییرات سلیقه کاربر در این روش مدل نمی‌شود. در خوشه‌بندی اخبار، با توجه به گوناگونی اخبار تعیین تعداد خوشه‌ها نیز مهم می‌باشد.

استفاده از تابع زمانی به عنوان یکی از ابزارهایی می‌باشد که در بهبود مدلسازی کاربر استفاده می‌شود. در این روش برای مدلسازی سلیقه کاربر علاوه بر ویژگی بلند مدت و کوتاه مدت از یک تابع زمانی نیز برای ارزش‌دهی به عناوین موجود در بردارهای ویژگی استفاده می‌شود. به این ترتیب عناوین با قدمت زمانی کمتر، داری وزن بیشتری می‌باشند [۷۴]. در این مقاله همچنین پس از اعمال سلیقه بلند مدت و بازیابی اخبار مرتبط با کاربر متقاضی، برای انتخاب اخبار متناسب با سلیقه کاربر از یک گراف همبستگی کاربر- خبر استفاده می‌شود. این گراف همبستگی رابطه بین اخبار (مرتبط با کاربر متقاضی) و کاربران را نشان می‌دهد. در این گراف با شروع از گره کاربر متقاضی، گره‌های خبری که متناسب با سلیقه کاربر باشند انتخاب می‌شوند. از آنجاییکه در این گراف از اخبار مورد علاقه کاربران مشابه نیز استفاده می‌شود، از اینرو پیشنهاد اخبار جدید به کاربر متقاضی نیز امکان‌پذیر می‌باشد.

### ج) سیستم‌های توصیه‌گر مبتنی بر اطلاعات فردی

تفاوت این الگوریتم نسبت به سایر روش‌ها در نحوه ایجاد ویژگی برای کاربر می‌باشد. در این روش برای ساخت ویژگی کاربر از اطلاعات شخصی(مثلا سن=۴۰)، رفتاری(مثلا موسیقی=کلاسیک)، و

اجتماعی(مثلا تحصیلات= کارشناسی) استفاده می‌شود. سپس به کمک این ویژگی کاربران دسته-بندی می‌شوند. با دسته‌بندی کاربران، در حقیقت سلایق مختلف شناسایی می‌شوند. سپس برای کاربر متقاضی، ابتدا براساس ویژگی وی، گروه کاربران مشابه شناسایی می‌شود و سپس منابع مورد استفاده این گروه به کاربر متقاضی پیشنهاد می‌شود.

در مقاله [۷۵] بر ایجاد دسته‌ها به صورت افزایش تاکید شده است. در این مقاله اگر یک کاربر به گروه‌های موجود، نزدیک نباشد آنگاه یک دسته جدید ایجاد می‌شود. همچنین در تکرارهای مختلف اطلاعات دسته‌ها براساس کاربر جدید تکمیل می‌شوند و یا در برخی موارد به کمک پرس‌وجوهایی اطلاعات اضافی از کاربر دریافت می‌گردد. یکی از مزایای این روش، عدم نیاز به سابقه کاربر برای توصیه می‌باشد. به عبارتی برای کاربر جدید نیز می‌تواند توصیه منبع را ارائه نماید. یکی از معایب این روش نحوه ایجاد ویژگی یا استخراج ویژگی‌های کاربران برای ساخت ویژگی می‌باشد. در این روش بر پرس‌وجو جهت دریافت داده‌ها تاکید شده‌است.

در مقاله [۷۶] ضمن بیان الگوریتم‌های هوش مصنوعی جهت دسته‌بندی، سیستم توصیه‌گری مبتنی بر اطلاعات آماری ( Demographics) برای گردشگری ایجاد شده‌است. در این مقاله برای جمع‌آوری داده‌های لازم جهت ساخت ویژگی کاربر از سیستم Tripadvisor.com استفاده شده‌است.

با توجه به محدودیت‌هایی که در جمع‌آوری اطلاعات در این حوزه وجود دارد، جهت کاهش وابستگی به کاربر جهت ثبت اطلاعات، از پردازش وب سایت شخصی نیز برای استخراج داده‌های لازم استفاده می‌شود [۷۷].

### د) سیستم‌های توصیه‌گر مبتنی بر شبکه‌های اجتماعی

این سیستم از داده‌های حاصل از گشت‌وگذار(ناوبری) در زمان گذشته استفاده می‌کند. این داده‌ها از ثبت تاریخچه گشت‌وگذار(ناوبری) توسط موتورهای جستجو، سرویس دهنده‌های وب، مرورگرها، صفحات وب، و لینک بین صفحات حاصل می‌شوند. این روش ضمن استخراج یک مدل گشت-

وگذار(ناوبری) از داده‌های موجود، در مورد بازدید صفحات در مراجعات بعدی یا توسط افراد جدید به کاربران کمک می‌کند.

در کارهای اولیه‌ای که در این زمینه انجام شده‌اند، ابتدا براساس جستجوهایی که کاربر انجام می-دهد، شبکه اجتماعی که کاربر متعلق به آن می‌باشد مدلسازی می‌گردد. این شبکه اجتماعی که می-تواند منابع(به عنوان مثال یک صفحه وب) و یا افراد خبره را مدل نماید، به مرور زمان به کمک پرس‌و جو و نتایج جستجوهای کاربر کامل می‌شود. سپس در مراجعات بعدی کاربر، براساس ساختار شبکه موجود، نتایج جستجو به کاربر نمایش داده می‌شوند. به عنوان مثال برای جستجو برای کلمه "annealing"، در حالت عادی ممکن است منابع متعددی پیشنهاد شود اما اگر در مدلسازی شبکه اجتماعی، کاربر متقاضی در شبکه‌ای با فیلد کامپیوتر قرار گیرد، منابع متعلق به "علوم کامپیوتر" به وی ارائه می‌شوند و موضوعات مربوط به "ذوب فلزات" به وی نمایش داده نمی‌شوند [۷۸, ۷۹].

ایده "دنباله فعالیت" برای شناسایی سلیقه کاربر به عنوان یکی از روش‌هایی می‌باشد که در این حوزه مطرح شده‌است. با استفاده از گشت‌وگذارهایی که کاربر در طول زمان در صفحات وب انجام داده، ویژگی از فعالیت کاربر ایجاد می‌شود. این ویژگی به " دنباله فعالیت" کاربر معروف می‌باشد. این دنباله فعالیت، دنباله‌ای از منابعی است که کاربر در تعاملات مکرر با مرورگر انتخاب نموده است. برای پیشنهاد منبع، دنباله فعالیت کاربر با دنباله فعالیت‌های ثبت شده مقایسه می‌شود. برای این مقایسه از تکنیک پنجره گذاری استفاده می‌شود. بدین ترتیب آخرین قطعه از فعالیت کاربر با هر قطعه از دنباله فعالیت‌ها مقایسه می‌شود و در صورت وجود اشتراک، این قطعه جدید به کاربر پیشنهاد می‌شود. این روش تاکید دارد که در پالایش مشارکتی ارتباطی بین امتیازدهی‌ها وجود ندارد و یا به عبارتی زمینه فعالیت- حداقل از نظر زمان- در نظر گرفته نمی‌شود که دنباله فعالیت، این مشکل را مرتفع می‌سازد. این روش‌ها سعی در کمک به کاربر جهت جستجوی منابع با کمترین پرس‌وجو دارند. از مزیت‌های این روش در فاز جمع‌آوری داده می‌باشد [۷۱, ۷۵].

با توسعه این روش و رشد شبکه‌های اجتماعی، نگرش جدید در سیستم‌های توصیه‌گر اجتماعی ایجاد شده‌است. شبکه‌های اجتماعی به عنوان ورودی جدیدی برای سیستم‌های توصیه‌گر پدیدار شدند که زمینه بهبود عملکرد این سیستم را فراهم نمودند. از طرفی سیستم‌های توصیه‌گر نیز زمینه مشارکت بیشتر در شبکه‌ای اجتماعی و توسعه آنرا فراهم نمودند.

در مقاله [۸۰] یک سیستم توصیه‌گر اجتماعی، به عنوان سیستمی که از روابط اجتماعی بین افراد نیز به عنوان ورودی استفاده می‌کند، تعریف شده‌است. به عبارتی همراه با ماتریس بازخورد کاربر در الگوریتم پالایش مشارکتی و یا محتوایی منابع در الگوریتم مبتنی بر محتوا، شبکه ارتباطی کاربران نیز مورد توجه می‌باشد. ایده اصلی این سیستم براین استوار است که "معمولا افرادی که در یک شبکه اجتماعی هستند نسبت به گروه‌های تصادفی دارای سلیقه مشابه‌تری می‌باشند". رابطه ۲-۸، الگوریتم کلی سیستم توصیه‌گر اجتماعی را بیان می‌کند.

$$Social\ Recommender\ System = Social\ Information\ Model + Basic\ CF \qquad (۸-۲)$$

براساس رابطه بیان شده در رابطه ۲-۸ ، به کمک روابط اجتماعی ابتدا آشنایی[1] افراد بررسی می‌شود و سپس برای گروه‌های آشنا، معیار شباهت (الگوریتم پالایش مشارکتی) اعمال می‌گردد.

به عنوان مثال برای پیشنهاد منبع به کاربر $u_x$، ابتدا شبکه اجتماعی کاربر $u_x$ بدست می‌آید. و سپس معیار شباهت برای کاربر $u_x$ و افرادی که در شبکه کاربر $u_x$ هستند اعمال می‌شود. در نهایت منابعی که کاربران شبیه به $u_x$ از شبکه اجتماعی این کاربر، مناسب ارزیابی کردند به کاربر $u_x$ پیشنهاد می‌شود. در مراجع [۸۱, ۸۲] اعتماد[2] به عنوان یکی از پارامترهای مهم جهت شکل‌گیری شبکه اجتماعی بیان شده‌است. به عنوان مثال می‌توان شبکه Epinions.com را اشاره نمود. پارامتر اعتماد به معیاری جهت تعیین میزان شباهت مورد استفاده قرار می‌گیرد.

[1] - Familiarity
[2] - Trust

در مقاله [۸۳] معماری و نحوه پیاده‌سازی یک سیستم پاسخ‌گویی سوال بیان شده‌است. این سیستم با دریافت یک سوال افراد خبره‌ای که می‌توانند به این سوال پاسخ دهند را پیشنهاد می‌کند. افراد خبره‌ای که توسط سیستم پیشنهاد می‌شوند دارای دو ویژگی می‌باشند. ویژگی اول مربوط به حوزه تخصصی افراد خبره می‌باشد که باید در حوزه سوال بیان شده باشند. ویژگی دوم بر میزان شباهت فرد خبره و متقاضی تاکید دارد. به عبارتی هر چه این شباهت بیشتر باشد، اعتماد متقاضی به پاسخ فرد خبره بیشتر می‌باشد. در این معماری شبکه اجتماعی و ویژگی کاربران به عنوان ابزارهای مهم در تعیین این معیار بیان شدند.

### ذ) سیستم‌های توصیه‌گر مبتنی بر زمینه تعامل

زمینه تعامل عبارت است از پارامترهای محیطی و فردی که در تعامل کاربر با سیستم موثر می‌باشند. به عنوان مثال زمان، انگیزه، مکان، دما. در غالب کارهایی انجام شده، تاکید بر رابطه بین کاربر و منبع می‌باشد و زمینه تعامل در نظر گرفته نمی‌شود. حال آنکه با توجه به تحقیقات انجام شده، معمولا کاربران در انجام تعامل با سیستم در شرایط مختلف، کارهای متفاوتی انجام می‌دهند [۲۳, ۶۳, ۸۴-۸۶]. به عنوان مثال در سیستم توصیه‌گر مربوط به تورهای مسافرتی، در نظر گرفتن فصل سال در توصیه مکان مسافرت مهم می‌باشد و یا در پیشنهاد یک رستوران، موقعیت جغرافیایی متقاضی مهم می‌باشد.

یکی از مسائل مهم در بکارگیری زمینه تعامل، نحوه مدلسازی و فرموله کردن آن برای سیستم توصیه‌گر می‌باشد. در این روش سیستم توصیه‌گر به صورت رابطه ۲-۹ مدل می‌شود.

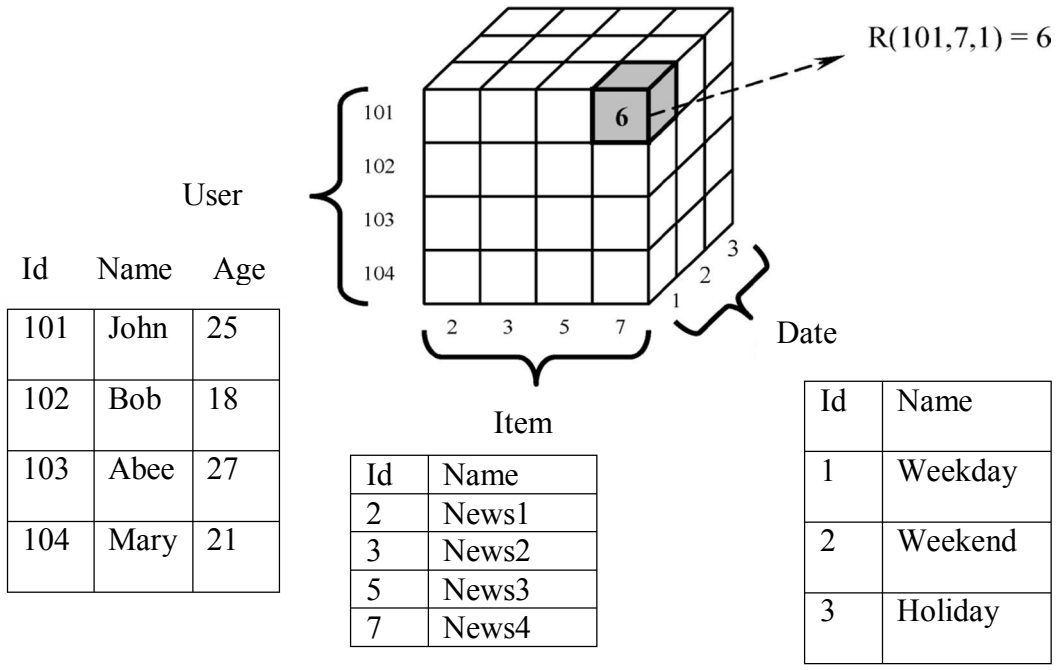
$$\begin{aligned} &R: U \times I \times C \rightarrow rate \\ &C = \{C^1, \dots, C^k\} \\ &C^1 = Date: \{Time, DayofWeek\}, \\ &C^2 = Location: \{Uni\ , Home\} \end{aligned} \tag{۲-۹}$$

در این رابطه U ویژگی کاربر و I ویژگی منبع و C زمینه تعامل را نشان می‌دهند. زمینه تعامل C خود می‌تواند شامل چند پارامتر مختلف باشد که هر یک دارای ویژگی خاصی باشند. به عنوان مثال در مشاهده خبر، زمان(Date) و مکان(Location) می‌توانند به عنوان دو پارامتر مستقل، زمینه تعامل کاربر با سیستم را تعریف نمایند. به این ترتیب برای سیستم توصیه‌گر خبر رابطه ۲-۹ را می‌توان به صورت رابطه ۲-۱۰ بازنویسی نمود. در نتیجه برای تعیین مقدار rate باید محاسبه چندبعدی یا چند پارامتری را مطابق رابطه ۲-۱۰ انجام داد.

$$R: U \times I \times Date \times Location \rightarrow rate \qquad (۱۰-۲)$$

در مقاله [۲۳] ایده پایگاه داده چند بعدی برای ذخیره داده‌ها و بازیابی موارد لازم استفاده شده-است. در این ایده ویژگی هر یک از مولفه‌هایی که زمینه تعامل را تعریف می‌کنند در جداول مجزا ذخیره می‌شوند و ارتباط هریک از مولفه‌ها به کمک جدول امتیازدهی معین می‌شود. به عنوان مثال شکل۲-۶، یک فضای سه بعدی برای نمایش ماتریس بازخورد خبر را نشان می‌دهد. با توجه به شکل می‌توان گزارش‌های تجمعی را نیز از ماتریس بازخورد استخراج نمود. رابطه ۲-۱۱ نمونه‌ای از این گزارش را نشان می‌دهد. این رابطه میانگین بازخوردها برای اخبار در زمان‌های پایان هفته از نظر کاربر $U_u$ را نشان می‌دهد.

$$R(U_u, Date_{weekend}) = AVG_{x.date=weekend}(U_u, x) \qquad (۱۱-۲)$$




شکل۲-۶: نمایش فضای سه بعدی از ماتریس بازخورد با درنظر گرفتن زمان به عنوان یک مولفه در زمینه تعامل [۲۳].

با توجه به شکل ۲-۶، یکی از مشکلات اساسی در این روش، نحوه محاسبه و پیش‌بینی امتیاز هر منبع می‌باشد. به همین دلیل برای بهبود حجم محاسبات، ابتدا ماتریس باخورد چند بعدی به ماتریس بازخورد دو بعدی($U \times I \rightarrow R$)تبدیل می‌شود و سپس روش‌های موجود برای پیش‌بینی امتیاز منبع جدید بکار می‌روند.

در مقاله [۸۵] نیز ضمن تاکید زمینه تعامل به عنوان یک عامل موثر در ارائه توصیه، روشی برای شناسایی ضمنی بازه فعالیت کاربر در تعامل با سیستم بیان شده‌است. در این روش زمینه تعامل در هر بازه به صورت ضمنی تعیین می‌شود. برای تعیین بازه فعالیت کاربر، دنباله تعاملات کاربر در طول زمان بررسی می‌شود و سپس به ازای مشاهده هر پیک یا نقطه عطف در دنباله تعاملات، این نقطه به عنوان نقطه تغییرات سلیقه کاربر و یا به عبارتی نقطه تغییر زمینه تعامل در نظر گرفته می‌شود. بازه بین هر دو نقطه تغییر، به عنوان زمینه تعامل انتخاب می‌گردد. برای پیش‌بینی و ارائه توصیه، فعالیت-

های انجام شده در بازه اخیر به عنوان، سلیقه جاری و یا زمینه تعامل جاری در نظر گرفته می‌شوند. البته در برخی کاربردها از بازخوردهای وزن‌دار بازه‌ها برای پیش‌بینی استفاده می‌گردد.

در مرجع [۸۷] از امکانات گوشی‌های موبایل برای تعیین زمینه تعامل استفاده می‌گردد. با توجه به قابلیت‌های گوشی‌های موبایل امروزی، به کمک ابزای‌هایی مانند GPS، gyroscope و microphone می‌توان اطلاعات مفیدی از زمینه تعامل استخراج نمود. با توجه به مطالعات انجام شده، نحوه جمع-آوری و مدلسازی و محاسبه داده‌های زمینه تعامل به عنوان مسائل اساسی در این حوزه می‌باشند.

### ر) سیستم‌های توصیه‌گر مبتنی بر پایگاه دانش

روش‌هایی که تا این بخش معرفی شده‌اند، از دانش انتزاعی بدست آمده از سلیقه کاربر استفاده می‌کنند. این دانش انتزاعی از ویژگی کاربر، ویژگی منبع، شبکه اجتماعی، بازخوردها، و زمینه تعامل استنتاج می‌شود. در روش مبتنی بر پایگاه دانش علاوه بر موارد فوق از یک پایگاه دانش مجزای دیگری نیز استفاده می‌شود. این پایگاه دانش به عنوان یک ورودی صریح به سیستم توصیه‌گر اعمال می‌شود. این پایگاه دانش حاوی اطلاعاتی در مورد منبع، نیاز یا سلیقه کاربر، و قوانین و محدویت‌هایی در مورد شرایط(زمینه) توصیه و ارتباط بین هر یک می‌باشد. در این دسته از روش‌ها به کمک این پایگاه دانش، رابطه بین کاربر و منبع مدل می‌شود [۸۸].

ایجاد پایگاه دانش مناسب، به عنوان یکی از مسائل مهم این روش می‌باشد. این دانش ممکن است به صورت قوانینی هنگام ایجاد سیستم توصیه‌گر به آن ارائه شود و یا به کمک ابزارهای وابسته به آنتالوژی استخراج شود. از مزایای این روش مقابله با مشکل کاربر/منبع جدید[1] و عدم نیاز به حجم داده‌ای زیاد نسبت به سایر روش‌ها می‌باشد. از طرفی اگر مشخصه منبع پیچیده باشد، وجود این پایگاه دانش نقش موثری در ارائه توصیه دارد. روش‌های مبتنی بر پایگاه دانش، از دو الگوریتم جستجو مبتنی بر نمونه و جستجو مبتنی بر قانون برای ارائه منابع پیشنهادی استفاده می‌کنند. در الگوریتم

[1] - Cold Start

مبتنی بر نمونه، براساس نمونه‌های قبلی که کاربر استفاده نموده و یا یک نمونه پیش فرض، سعی می‌شود، سایر منابع مشابه بازیابی و به کاربر نمایش داده شوند. در فرایند جستجو ممکن است، بازخوردهایی از کاربر دریافت شود و از این بازخوردها برای ادامه جستجو استفاده نماید[۸۹].

در مقاله [۸۹] سیستم توصیه‌گری برای اجاره منزل ارائه شده‌است. در این سیستم ابتدا یک پایگاه دانش اولیه از منازل ایجاد شده‌است. به ازای درخواست کاربر، ابتدا نیاز کاربر به پایگاه دانش اضافه می‌شود و سپس براساس دانش بدست آمده از کاربر و منازل موجود، پیشنهاد منزل انجام می‌شود. به عنوان مثال اگر کاربر برای اولین بار با درخواست "منزلی با قیمت x و تعداد اتاق n و در مکان z " به سیستم مراجعه نماید. ابتدا نزدیک‌ترین نمونه به کاربر پیشنهاد می‌شود و سپس بازخورد کاربر مبنی بر اصلاح ویژگی مورد پیشنهاد شده دریافت می‌گردد. سپس مجددا نمونه‌های جدید جستجو و پیشنهاد می‌شوند. بازخورد نهایی کاربر به عنوان دانشی(ویژگی) از کاربر نگهداری می‌شود.

در الگوریتم مبتنی بر قانون، علاوه بر تعریف منبع، نیاز کاربر و محدودیت‌های استفاده از منبع و شرایط محیط به صورت قوانین در پایگاه دانش ایجاد می‌شوند. در مقاله [۹۰] سیستم توصیه‌گر مبتنی بر پایگاه دانش به کمک الگوریتم جستجوی مبتنی بر قانون برای سرویس میزبان وب ارائه شده‌است. پایگاه دانش شامل ویژگی کاربر متقاضی ( نظیر تعداد کاربر همزمان، قیمت، سرعت) و ویژگی سرویس‌ها ( نظیر قیمت، حجم، پهنای باند) و محدودیت‌های منابع (میزبان وب با تعداد کاربر بیش از ۲۵۰ در روز نیاز به سرعت بالا دارد) و محدودیت درخواستی کاربر (که رابطه بین کاربر و منبع را تعیین می‌کند، مثلا قیمت سرویس کمتر از x باشد) می‌باشد. به این ترتیب سیستم توصیه‌گر با دریافت ویژگی کاربر متقاضی، ویژگی منابع، محدودیت‌ها، پردازش لازم برای پیشنهاد منبع مناسب را انجام می‌دهد. به عبارتی منابعی را که با محدودیت‌ها و ویژگی کاربران و منابع سازگار باشند را به کاربر متقاضی پیشنهاد می‌نماید.

### ز) سیستم‌های توصیه‌گر ترکیبی

روش‌هایی که تاکنون بیان شده‌اند، هر یک دارای معایب و مزایایی می‌باشند که می‌توانند در کنار یکدیگر تا حدی همپوشانی نمایند و مشکلات را کاهش دهند. از این‌رو، ترکیب این روش‌ها و ارائه سیستم توصیه‌گر جدید به عنوان یک راهکار مهم در جهت بهبود عملکرد سیستم‌های توصیه‌گر می‌باشد. اما انتخاب الگوریتم‌ها و نحوه ترکیب آنها در بهبود عملکرد سیستم توصیه‌گر بسیار مهم می‌باشد. مکمل بودن الگوریتم‌ها یکی از نکات مهمی است که باید در انتخاب الگوریتم‌ها دقت شود.

به عنوان نمونه، سیستم‌های توصیه‌گر مبتنی بر دانش مشکل کاربر/منبع جدید ندارند. یعنی برای کاربر جدید هم امکان پیش‌بینی منبع وجود دارد. زیرا دانش منابع از قبل وجود دارد و قوانین انتخاب نیز به همراه درخواست کاربر جدید ارائه می‌شود و نیاز به دانش قبلی نمی‌باشد. اما سیستم‌های مشارکتی نیاز به یک ویژگی مناسب از کاربر دارد تا بتواند پیش‌بینی مناسبی انجام دهد که امر نیاز به زمان می‌باشد. در روش ارائه شده، در ابتدا که ویژگی تعامل و بازخورد کاربر کم می‌باشد از سیستم مبتنی بر دانش استفاده می‌شود. با تهیه ویژگی مناسب از کاربر، در تعاملات بعدی، نتایج حاصل از سیستم توصیه‌گر مشارکتی نیز لحاظ می‌شود. البته بحث هزینه محاسباتی از معایب این روش می‌باشد.

به عنوان نمونه در مقاله [۳۶] یک مدل ترکیبی با مولفه‌های پالایش مشارکتی مبتنی بر حافظه و پالایش مشارکتی مبتنی بر مدل برای توصیه خبر ارائه شده‌است. در این روش بازخورد هر خبر به کمک "انجام کلیک" تعیین می‌شود. در مولفه مبتنی بر مدل با استفاده از ماتریس بازخورد بدست آمده ابتدا کاربران خوشه‌بندی می‌شوند. هر کاربر ممکن است بر اساس درجه عضویت به چند خوشه متعلق باشد. برای تعیین محبوبیت خبر s از نظر کاربر u از رابطه ۲-۱۲ استفاده می‌گردد:

$$r_{u,s} = \sum_{ci:u \in ci} w(u, ci) * \sum_{uj:uj \in ci} I(uj, s) \qquad (۱۲-۲)$$

در این رابطه w(u, ci) درجه عضویت کاربر u نسبت به خوشه ci را نشان می‌دهد. I(uj ,s) بیان-کننده کلیک شدن یا نشدن خبر s توسط کاربر uj می‌باشد. در مولفه مبتنی بر حافظه، از شباهت خبر s با اخبار مطالعه شده توسط کاربر u استفاده می‌گردد. در پایان برای امتیاز دهی به خبر و ارائه توصیه از ترکیب وزنی این دو مولفه استفاده می‌گردد. در مقاله [۶۲] نیز یک سیستم توصیه‌گر ترکیبی ارائه شده‌است که از دو مولفه پالایش مشارکتی مبتنی بر منبع و تئوری بیز استفاده می‌کند. در این سیستم ابتدا براساس الگوریتم بیز، پیش‌بینی انجام می‌شود. اگر این پیش‌بینی از حدآستانه کمتر باشد، الگوریتم مشارکتی مبتنی بر منبع استفاده می‌شود در غیر اینصورت نتیجه بیز برای پیشنهاد منبع استفاده می‌گردد. از آنجاییکه روش بیز به صورت غیر همزمان(آفلاین) انجام می‌شود از اینرو نسبت به الگوریتم مشارکتی مبتنی بر منبع هزینه محاسباتی کمتری دارد. البته هرگاه الگوریتم بیز نتواند پیش‌بینی مناسبی انجام دهد، از الگوریتم مشارکتی مبتنی بر منبع استفاده می‌گردد که می‌تواند نتیجه پیش‌بینی را بهبود دهد.

ترکیب دو روش محتوایی و مشارکتی، به عنوان یکی از روش‌های معمول برای سیستم‌های توصیه-گر ترکیبی می‌باشد. به کمک الگوریتم محتوایی، ویژگی محتوایی برای هر کاربر ایجاد می‌شود. این ویژگی شامل منابعی است که کاربران استفاده کرده‌اند. در این روش ابتدا به کمک ویژگی محتوایی، کاربران مشابه بازیابی می‌شوند و سپس به کمک الگوریتم مشارکتی امتیازدهی منابعی که مطابق ویژگی کاربر متقاضی می‌باشند انجام و لیست مرتب شده‌ای از منابع ارائه می‌شود [۹۱].

مولفان در مقاله [۹۲] از ویژگی خوشه‌بندی شده کاربران برای خوشه‌بندی اخبار ورودی و پیشنهاد اخبار جدید استفاده می‌کنند. در این روش ابتدا ماتریس کاربر-کلمه ایجاد می‌شود که هر درایه آن وزن کلمه در ویژگی کاربر نظیر را نشان می‌دهد که با استفاده از مدل LDA[1] [۹۳] محاسبه می‌شود. با استفاده از این ماتریس، کاربران خوشه‌بندی می‌شوند. ویژگی کاربران متعلق به یک خوشه با هم ترکیب می‌شوند. با استفاده از ویژگی خوشه‌بندی شده کاربر، اخبار جدید دو به دو بررسی می‌شوند که

[1] - Latent Dirichlet Allocation

آیا متعلق به دسته مورد نظر می‌باشند یاخیر. در نتیجه این مقایسه‌ها، به ازای هر خوشه ماتریس‌هایی تولید می‌شوند که درایه‌های هر ماتریس میزان تعلق دو خبر نظیر به خوشه مورد نظر را نشان می‌دهند. سپس با استفاده از ماتریس‌های موجود و تکنیک ensemble hierarchical clustering، خوشه‌بندی اخبار انجام می‌شود. پس از خوشه‌بندی اخبار ابتدا خوشه‌های مطابق با سلیقه کاربر شناسایی و اخبار این خوشه‌ها جمع‌آوری می‌شوند. سپس اخبار متناسب با سلیقه کاربر از بین این اخبار انتخاب می‌شوند.

از مسائل مهم در توسعه سیستم‌های توصیه‌گر ترکیبی، نحوه ترکیب مولفه‌ها می‌باشدکه چند مورد از آنها در ادامه بیان شده‌اند [۲۵, ۹۴].

- ترکیب وزنی: در این روش برای بدست آوردن امتیاز پیش‌بینی شده منبع توسط سیستم، امتیاز پیش‌بینی شده توسط هریک از مولفه‌ها به صورت ترکیب خطی وزن دار، جمع می‌شوند و امتیاز نهایی منبع را بدست می‌دهند.

- انتخابی: در این روش، یک حد آستانه تعیین می‌شود و معمولا خروجی یکی از مولفه‌ها به عنوان پیش فرض انتخاب می‌گردد. اگر خروجی این مولفه از حد آستانه کمتر باشد، خروجی مولفه دوم انتخاب می‌گردد.

- ترکیبی: در این روش، نتایج حاصل از هر دو الگوریتم در خروجی به کاربر نشان داده می‌شوند. البته این روش برای کاربردهایی مناسب است که تعداد نتایج زیاد در خروجی مطلوب باشد.

- سلسله مراتبی: در این روش خروجی یک مولفه به عنوان ورودی مولفه دیگر استفاده می‌گردد.

هم‌افزایی ویژگی‌ها[1]: در این روش ابتدا با کمک روش محتوایی داده‌های موجود به سیستم آموزش داده می‌شوند و سپس از مدل ایجاد شده برای تخمین امتیاز منابعی که بازخورد ندارند استفاده می‌شود. سپس به کمک داده‌های جدید، پیش‌بینی انجام می‌شود. در حقیقت در این روش ابتدا درایه‌های

[1] - Feature Augmentation

ماتریس بازخورد تخمین زده می‌شوند و سپس (به عنوان مثال) الگوریتم مشارکتی بر روی ماتریس بازخورد جدید اعمال می‌گردد.

## ۲-۳- مروری بر رویکرد مدلسازی رفتار کاربر در توسعه سیستم‌های توصیه‌گر

صرفنظر از الگوریتم مورد استفاده، سیستم‌های توصیه نیازمند ایجاد یک مدل مناسب از پروفایل کاربر جهت بکارگیری در مکانیسم ارائه توصیه و پیشنهاد منابع متناسب با نیاز (سلیقه) می‌باشند. ساخت این مدل مناسب نقش اساسی در موفقیت سیستم توصیه‌گر دارد [۲۸, ۹۵]. همانطور در بخش ۲-۱ بیان شد، مدلسازی ساختاری و مدلسازی رفتاری، دو رویکرد اصلی در نحوه بکارگیری پروفایل کاربر در الگوریتم‌های توصیه می‌باشند.

**رویکرد ساختاری**. در این رویکرد پروفایل کاربر با استفاده از یک بردار ویژگی ( ساختار تجریدی) مدل می‌شود. یکی از روش‌های معمول برای ایجاد مدل ساختاری استفاده از ساختار مجرد بصورت بردار ویژگی $< userId, itemId, feedback, timestamp >$ می‌باشد. در پالایش مشارکتی مبتنی بر شباهت، بردارهای ویژگی کاربران برای ساخت ماتریس بازخورد استفاده می‌شود. این ماتریس بازخورد برای یافتن شباهت بین کاربران یا منابع استفاده می‌شود. در پالایش مشارکتی مبتنی بر مدل، تکنیک‌های یادگیری ماشین نظیر k-means، Bayse، و support vector machine بر ماتریس بازخورد اعمال و از مدل ایجاد شده برای پیش‌بینی منابع استفاده می‌شود [۱, ۲]. همچنین در پالایش مشارکتی مبتنی بر فاکتورهای پنهان، با اعمال تکنیک تجزیه ماتریس (بر ماتریس بازخورد)، سلیقه کاربر استخراج و از آن در پیش‌بینی امتیاز منابع استفاده می‌شود [۳۳, ۳۷, ۴۷]. در پالایش محتوایی، مدل ساختاری کاربر با استفاده از بردار ویژگی $< userId,\ item - feature - vector, timestamp >$ ساخته می‌شود. بدین ترتیب ویژگی منابع (نظیر کارگردان یک فیلم یا سناریوی یک فیلم) انتخاب شده توسط کاربر در بردار ویژگی استفاده می‌شود [۲۲].

**رویکرد رفتاری.** در این رویکرد، پارامترهای موثر بر رفتار کاربر در انتخاب منابع، با استفاده از پروفایل کاربر استخراج می‌شود. در این روش برخی از ویژگی پروفایل کاربر نظیر زمان، بازخورد (امتیاز) به عنوان مشاهدات (پارامترهایی با مقدار معین) و برخی دیگر نظیر سلیقه کاربر نیز به عنوان فاکتورهای پنهان در نظر گرفته می‌شوند. استفاده از چارچوب احتمالاتی گرافیکی نظیر شبکه بیزین، یکی از روش‌های معروف در ایجاد مدل رفتاری می‌باشد. مولفان در [۳۰] یک مدل گرافیکی برای پیشنهاد منابع معرفی کرده‌اند. در این مدل گرافیکی سلیقه کاربر با استفاده از پارامترهای پنهان استخراج می‌شود. سپس با استخراج گرایش‌های کلی (زمینه تعامل) از منابع، منابع مشابه با سلیقه هر کاربر به وی پیشنهاد می‌شود. در [۹۶] نیز از چارچوب شبکه بیزین برای ایجاد یک سیستم توصیه‌گر استفاده شده‌است. این روش که برای پیشنهاد اطلاعات در یک گروه اجتماعی معرفی شده‌است، از اطلاعات محتوایی منابع انتحاب شده توسط کاربر (رویکرد پالایش محتوایی) برای شناسایی سلیقه کاربر و پیشنهاد منابع بهره می‌برد. مدل مارکوف نیز به عنوان یکی دیگر از چاچوب‌های ایجاد مدل رفتاری مورد استفاده قرار می‌گیرد. در [۳۵]، ماتریس بازخورد در یک مدل مارکوف مخفی بکارگرفته شده‌است. در این روش با رویکرد پالایش مشارکتی، براساس منابعی که کار در طول زمان انتخاب نموده‌است، سلیقه کاربر را در زمان جدید تخمین می‌زند.

## ۲-۴-ارزیابی سیستم‌های توصیه‌گر

با توجه به اینکه توسعه سیستم‌های توصیه‌گر به رضایت کاربران بستگی دارد، بنابراین وجود معیارهایی که بتوان به کمک آنها عملکرد سیستم‌های توصیه‌گر را ارزیابی نمود، ضروری می‌باشد. برای ارزیابی سیستم‌های توصیه‌گر از سه روش تست استفاده می‌شود [۹۷, ۹۸] و در هریک از این روش‌ها نیز معیارهایی برای ارزیابی بکار می‌روند. در ادامه ابتدا روش‌های تست معرفی می‌شوند، سپس معیارهای ارزیابی بیان می‌شوند.

- روش ارزیابی مبتنی بر کاربر فعال:

در این روش سیستم توصیه‌گر در محیط واقعی توسط کاربران مورد استفاده قرار می‌گیرد. در این روش علاوه بر تست کارائی سیستم می‌توان رضایت‌مندی کاربران را نیز در فضای واقعی بررسی نمود. البته انتخاب گروه‌های کاربری مناسب جهت حصول نتیجه قابل اعتماد اهمیت زیادی دارد و مهمتر اینکه ممکن است ارائه نتایج نامناسب استقبال کاربران را به استفاده مجدد از سیستم کاهش دهد.

- روش ارزیابی مبتنی بر کاربر کنترل شده (فعال کنترل شده):

در این روش تعدادی کاربر برای انجام تست انتخاب می‌شوند. این کاربران از فرایند تست با خبر می‌باشند و گاهی نیز نحوه تعامل با سیستم جهت ارزیابی مناسب به کاربران توصیه می‌شود. به دلیل رصد کاربر، امکان اخذ گزارش‌های متعدد برای سیستم را فراهم می‌کند. از طرفی تهیه گزارش‌های کیفی نیز از طریق کاربران ممکن می‌باشد. به علت محدود بودن کاربران امکان تهیه داده‌های آماری زیاد فراهم نمی‌باشد. تاکید این روش بر تست عملکردی سیستم می‌باشد.

- روش ارزیابی مبتنی بر پایگاه داده‌ها (غیرفعال):

در این روش به کمک داده‌هایی که از تعاملات گذشته کاربران بدست آمده جهت ارزیابی سیستم استفاده می‌گردد. از جمله منابع داده‌ای که عموما (بویژه در روش‌های پالایش مشارکتی) مورد استفاده قرار می‌گیرد، مجموعه داده‌ای MovieLens[1] می‌باشد. از جمله مزیت‌های این روش امکان تکرارپذیری آزمایش در شرایط مشابه و کنترل پارامترهای آزمایش می‌باشد. از جمله مشکلات این روش نیز نحوه مدل کردن رفتار کاربر می‌باشد بنحوی که بتواند شرایط تعامل واقعی را نشان دهد. البته با استفاده از مشخصه زمان در داده‌ها می‌توان ترتیب انجام تعاملات و زمان آنها را رصد نمود.

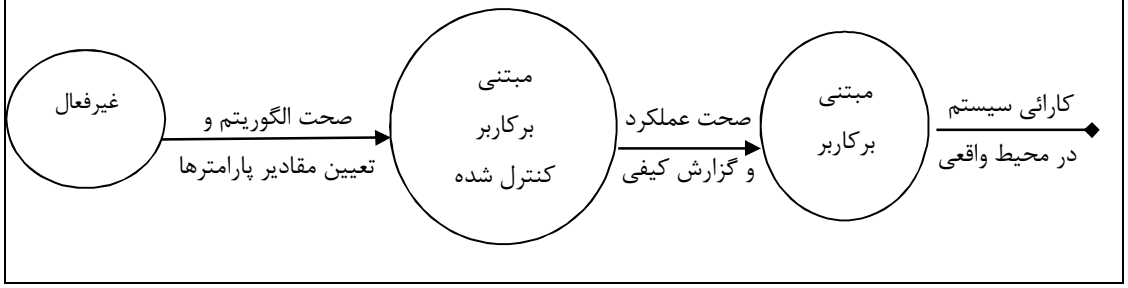
جدول۲-۲ مقایسه‌ای از روش‌های بیان شده جهت ارزیابی سیستم توصیه‌گر را نشان می‌دهد. برای استفاده عملی از این روش‌ها، معمولا ابتدا روش غیرفعال استفاده می‌گردد که در نتیجه آن صحت الگوریتم و تست پارامترها انجام می‌شود، سپس روش مبتنی برکاربر کنترل شده استفاده می‌گردد که

[1] - https://movielens.org/

در نتیجه آن صحت عملکرد و گزارش‌های کیفی حاصل می‌شوند و در نهایت روش مبتنی بر کاربر جهت بررسی کارائی سیستم در محیط واقعی بررسی می‌شود. البته در تحقیقات انجام شده، غالبا روش ارزیابی مبتنی بر پایگاه داده استفاده شد. روش‌های مختلفی که برای ارزیابی سیستم‌های توصیه‌گر استفاده می‌شوند، معیارهایی را برای ارزیابی بکار می‌گیرند که در ادامه این معیارها و نحوه محاسبه آنها بیان می‌شوند.

جدول۲-۲: مقایسه روش‌های مختلف ارزیابی سیستم‌های توصیه‌گر.

| روش ارزیابی/ ویژگی | مزیت | عیب |
|---|---|---|
| مبتنی بر کاربر | بررسی کارائی سیستم در محیط واقعی | عدم استقبال کاربران به علت برخی پیشنهادهای نادرست در حین ارزیابی |
| مبتنی بر کاربر کنترل شده | امکان دریافت گزارش‌های متعدد کیفی و کمی | تعداد کم داده تستی و همگرایی کاربران به جلب رضایت مدیر سیستم |
| غیرفعال(آفلاین) | تکرارپذیری آزمایش در شرایط ثابت و نیز در صورت تغییر پارامترها | مدل کردن کاربر و یا به عبارتی شبیه‌سازی رفتار کاربر |




- سازگاری با سلیقه کاربر:

یکی از ویژگی‌های مهم سیستم‌های توصیه‌گر میزان سازگاری آن با سلیقه کاربر می‌باشد. معمولا برای ارزیابی این ویژگی از روش ارزیابی مبتنی بر کاربر کنترل شده استفاده می‌شود. زیرا در این روش، ارزیابی کیفی از طریق پرس‌وجو با کاربر امکان پذیر می‌باشد.

- دقت پیش‌بینی:

منظور از دقت پیش‌بینی، تناسب میزان امتیاز ارزیابی شده برای یک منبع نسبت به امتیاز مورد نظر کاربر می‌باشد. به عنوان مثال اگر کاربر امتیاز یک منبع را $a$ ارزیابی نمود، سیستم توصیه‌گر خوب نیز باید امتیاز $a \pm \varepsilon$ را به این منبع نسبت دهد. برای ارزیابی میزان صحت از سه معیار استفاده شده که در ادامه بیان می‌شوند.

- تعیین دقت امتیاز پیش‌بینی شده:

برای سیستم‌هایی که به کمک مقادیر کمی یا کیفی رتبه‌بندی می‌شوند می‌توان از این روش استفاده نمود. در این روش از میانگین خطا و مجذور میانگین خطا استفاده می‌گردد. روابط ۲-۱۳ و ۲-۱۴ به ترتیب نحوه محاسبه میانگین خطا و میانگین مجذور خطا را نشان می‌دهند.

$$RMAE = \sqrt{\frac{1}{|S|}\sum_{u_i \in S} |\hat{r}_{u_i} - r_{u_i}|} \qquad (۱۳-۲)$$

$$RMSE = \sqrt{\frac{1}{|S|}\sum_{u_i \in S} (\hat{r}_{u_i} - r_{u_i})^2} \qquad (۱۴-۲)$$

- ارزیابی مفید بودن منابع پیش‌بینی شده:

در این روش به جای تاکید بر درست بودن امتیاز پیش‌بینی شده برای منبع پیشنهادی، بر مفید بودن منبع پیشنهادی تاکید می‌شود. در جدول۲-۳، حالت‌های مختلف مفید بودن و مفید نبودن منبع، و نیز انتخاب یا عدم انتخاب هر یک آمده‌است.

جدول۲-۳: حالت های مختلف مفید بودن /نبودن منبع و انتخاب هریک از منابع.

| | منابع توصیه شده | منابع توصیه نشده |
|---|---|---|
| منابع مفید (استفاده شده) | منابع مفید توصیه شده True-Positive(TP) | منابع مفید توصیه نشده False-Negative(FN) |
| منابع غیرمفید (استفاده نشده) | منابع غیر مفید توصیه شده False-Positive(FP) | منابع غیرمفید توصیه نشده True-Negative(TN) |

با توجه به جدول۲-۳، می‌توان معیارهای "درصد پوشش منابع مفیدی که سیستم می‌تواند توصیه نماید"[1] و "درصد منابع مفید توصیه شده"[2] و "درصد منابع غیرمفید توصیه شده"[3] را محاسبه نمود. روابط ۲-۱۵ و ۲-۱۶ و ۲-۱۷، به ترتیب نحوه محاسبه این معیارها را نشان می‌دهند.

$$Recall = \frac{TP}{TP + FN} \tag{۱۵-۲}$$

$$Precision = \frac{TP}{TP + FP} \tag{۱۶-۲}$$

$$Fallout = \frac{FN}{TN + FN} \tag{۱۷-۲}$$

معیار دیگری نیز معرفی شده‌است که به $\mathrm{F-Measure}(\mathrm{F}_\beta)$ معروف می‌باشد که و کمک مقادیر Precision و Recall محاسبه می‌شود. این معیار نیز جهت ارزیابی میزان کارائی سیستم توصیه‌گر استفاده می‌شود. رابطه ۲-۱۸، نحوه محاسبه فرم معمول این معیار(به ازای $\beta = 1$) را نشان می‌دهد.

$$F1 = \frac{2*Precision*Recall}{Precision*Recall} \tag{۱۸-۲}$$

[1] - Recall
[2] - Precision
[3] - Fallout

- مطلوبیت لیست منابع پیشنهاد شده:

هنگامی که خروجی یک سیستم توصیه‌گر همانند موتورهای جستجو یک لیست پیشنهادی باشد، استفاده از رتبه منبع توصیه شده در لیست پیشنهادی می‌تواند جهت ارزیابی کارائی سیستم توصیه‌گر استفاده شود. در این روش سیستم توصیه‌گر خوب باید منابع مفیدتر را در ابتدای لیست پیشنهاد نماید. برای تعیین این معیار دو حالت در نظر گرفته می‌شود. در حالت اول، یک لیست مطلوب کاربر از قبل وجود دارد که به لیست مرجع معروف است. در این حالت میزان خطای لیست پیشنهادی نسبت به لیست مرجع محاسبه می‌شود. در حالت دوم - که معمولا در روش‌های مبتنی‌بر کاربر مطلوب است- رتبه منبع در لیست با امتیازی که کاربر به آن می‌دهد (ترتیب انتخاب) مقایسه می-شود.

- میزان پوشش داده‌های ورودی[1]:

وجود داده‌ای کافی در ارائه نتایج بهتر و صحت نتایج یک سیستم توصیه‌گر ضروری می‌باشد. اما افزایش حجم داده (صرفنظر از داده نامرتبط) سبب کاهش سرعت پاسخ‌گویی سیستم می‌شود. بنابراین ممکن است یک سیستم توصیه‌گر از بین N داده فقط n داده را استفاده نماید که $n<<N$ می‌باشد. برای ارزیابی این معیار می‌توان منابعی که در یک آزمایش استفاده می‌شوند را نسبت به کل منابع محاسبه نمود. استفاده از این معیار در روش تست غیرفعال مطلوبیت بیشتری دارد.

- نحوه برخورد با داده(کاربر/منبع) جدید[2]:

یکی از معیارهای ارزیابی سیستم توصیه‌گر، توانایی ارائه توصیه به کاربران یا منابع جدید می‌باشد. یعنی آیا یک الگوریتم استفاده شده در سیستم توصیه‌گر می‌تواند برای کاربر جدید منبعی را پیشنهاد نماید، و از طرفی آیا قادر می‌باشد یک منبع جدید را در محاسبات بکار گیرد.

---

[1] - Coverage
[2] - Cold Start

- اطمینان سیستم به منابع پیشنهادی[1]:

در محاسبه امتیاز منابع و رتبه‌بندی آنها، سیستم می‌تواند یک ضریب اطمینان محاسباتی را به منابع پیشنهاد شده نسبت دهد. این ضریب بیان کننده میزان اعتماد سیستم به رتبه‌بندی یا امتیاز محاسبه شده می‌باشد. به عنوان مثال ممکن است منبع Si از بین ۱۰ رای، امتیاز ۴.۵ از ۵ را کسب نماید، و منبع Sj از بین ۱۰۰ رای، امتیاز ۴ از ۵ را کسب نماید. در این حالت اطمینان سیستم به امتیاز محاسبه شده برای منبع Sj بیشتر می‌باشد.

- اعتماد کاربران به سیستم[2]:

اعتماد کاربران به منابع پیشنهادی سیستم توصیه‌گر در توسعه آن سیستم تاثیر زیادی دارد. این اعتماد از طریق تجربه تعامل با سیستم و نیز ارائه توضیح در مورد فرایند(دلیل) پیشنهاد منابع و ارائه (ضریب)اعتماد سیستم به ارزیابی ارائه شده حاصل می‌گردد.

- میزان منابع جدید در توصیه‌ها[3]:

یکی از نیازهایی که سبب استفاده از سیستم توصیه‌گر می‌باشد، مشاهده منابع جدید است. یعنی سیستم بتواند منابعی که کاربر تا به حال مشاهده نکرده را بازیابی نماید. این امر در سیستم‌های تجاری اهیمت زیادی دارد. زیرا سبب افزایش فروش اقلام جدید می‌شود. برای ارزیابی این معیار می‌توان نسبت منابع جدید پیشنهاد شده به کل منابع پیشنهاد شده را محاسبه نمود. رابطه ۲-۱۹، نحوه محاسبه آنرا نشان می‌دهد.

$$Novelty = \frac{\text{تعداد منابع جدید توصیه شده}}{\text{کل منابع توصیه شده}} \qquad (۱۹-۲)$$

[1] - Confidence
[2] - Trust
[3] - Novelty

- قابلیت پیش‌بینی سلیقه کاربر[1]:

عبارت است از قابلیت سیستم توصیه‌گر در ارائه منابع جدیدی که سبب تغییر سلیقه کاربر گردد. یعنی بتوان سلیقه کاربر را پیش‌بینی نمود یا آنرا تغییر داد.

- گوناگونی در لیست پیشنهادی[2]:

این ویژگی به وجود منابع مختلف در لیست پیشنهادی تاکید دارد. مثلا برای یک سیستم توصیه‌گر مسافرت، مکان‌هایی که برای مسافرت پیشنهاد می‌شوند همگی در یک محیط جغرافیایی و با اختلاف در هتل نباشند. برای تعیین این ویژگی ابتدا باید یک معیار برای تعیین اختلاف(شباهت) منابع تعریف نمود و به کمک آن میزان گوناگونی سیستم را ارزیابی نمود.

- حفظ حریم خصوصی[3]:

کاربرانی که از یک سیستم توصیه‌گر استفاده می‌کنند انتظار دارند که اطلاعات شخصی آنها به صورت محرمانه نگهداری شوند و در اختیار دیگران قرار نگیرند و یا قابل بازیابی و استنتاج نباشند. این موضوع ممکن است در برخی موارد سبب کاهش صحت سیستم توصیه‌گر شود. به عنوان مثال در یک سیستم توصیه‌گر توریستی، برای پیشنهاد یک رستوران، تعیین مکان کاربر ضروری می‌باشد.

- توسعه (مقیاس) پذیری[4]:

توسعه‌پذیری و سازگاری با محیط‌های بزرگ از ویژگی‌های مهم سیستم‌های توصیه‌گر می‌باشد. زیرا این سیستم‌ها باید در سیستم‌های تجاری بزرگ که شامل داده و کاربران و تعداد تعاملات زیادی می‌باشند مورد استفاده قرار گیرند. برای ارزیابی این معیار، از اندازه‌گیری میزان منابع سیستمی مصرف شده، مدت زمان پاسخ‌گویی و تعداد پاسخ در واحد زمان (به ازای بکارگیری سیستم توصیه‌گر با حجم داده‌ای بزرگ) استفاده می‌شود.

---

[1] - Serendipity
[2] - Diversity
[3] - Privacy
[4] - Scalabilty

- سازگاری با سلیقه کاربر[1]:

یک سیستم توصیه‌گر باید بتواند به مرور زمان با دریافت تعداد بازخوردهای بیشتر خود را به سلیقه کاربر نزدیک نماید. نتیجه این سازگاری ارائه منابع مناسب به کاربر و افزایش استقبال کاربر به استفاده مجدد از سیستم توصیه‌گر می‌شود. در روش غیرفعال(آفلاین) برای ارزیابی این معیار می‌توان ابتدا با N داده سیستم توصیه‌گر را تست نمود و سپس با N+K آنرا تست نمود. برای مقایسه باید در هر حالت کارائی سیستم( مثلا معیار دقت) را ارزیابی نمود.

- مقابله با حملات[2]:

از آنجایی‌که سیستم‌های توصیه‌گر در محیط تجاری برای افزایش فروش اهمیت فراوانی دارند، از اینرو امکان حمله به این سیستم‌های توصیه‌گر جهت کاهش استقبال کاربران به سیستم توصیه‌گر و یا افزایش استقبال کاربران به یک منبع خاص وجود دارد. به عنوان مثال در سیستم توصیه‌گر کتاب، با تزریق امتیازهای غیر واقعی به کتاب، احتمال حضور این کتاب در لیست پیشنهادی به کاربران افزایش می‌یابد.

- میزان مطلوبیت منابع پیشنهادی[3]:

این ویژگی میزان سود و یا زیان حاصل از بکارگیری منابع پیشنهادی را ارزیابی می‌کند. هر چند امتیاز منابع بازیابی شده(یا ترکیب آن با میزان اطمینان سیستم) می‌تواند به عنوان یک معیار مطلوبیت مورد استفاده قرار گیرد، اما تعریف یک تابع مطلوبیت از مشکلات این روش می‌باشد.

یکی از مسائلی که باید در ارزیابی‌ها مورد توجه قرار گیرد، صحت ارزیابی می‌باشد. مواردی مانند انتخاب شانسی و یا بایاس بودن داده‌ها ممکن است منجر به عدم ارزیابی مناسب شوند. یکی از راه حل‌هایی که برای مقابله با این مشکل مطرح شده‌است، تکرار فرایند ارزیابی می‌باشد. در این روش یک الگوریتم چندین بار تست می‌شود تا ارزیابی بدست آمده قابل اعتماد باشد.

---

[1] - Adaptivity
[2] - Robustness
[3] - Utility

## ۲-۵- جمع‌بندی

در این فصل ضمن بیان روش‌های موجود برای توسعه سیستم‌های توصیه‌گر، دسته‌بندی از این روش‌ها انجام شده‌است. با توجه به روشهای موجود، حوزه‌های کاری که داده‌های مناسب برای سیستم‌های توصیه‌گر را فراهم می‌کنند نیز در مکانیسم ارائه توصیه موثر می‌باشند. همچنین، ایجاد مدل مناسب از کاربر جهت تخمین سلیقه آینده و بکارگیری الگوریتم مناسب جهت انجام پردازشهای لازم به عنوان چالشهای مهم این حوزه کاری می‌باشند.

با توجه به مطالعات انجام شده، غالب روش‌های موجود بر بکارگیری مدل ساختاری در یک مکانیسم توصیه مناسب جهت استخراج سلیقه کاربر تمرکز دارند. بدین ترتیب در این حوزه کاری، جهت‌گیری کلی براساس دو مسئله مدل کاربر و الگوریتم مناسب انجام می‌شود. در این رساله نیز بر ایجاد مدل مناسب از کاربر تاکید شده‌است که جزئیات آن در فصل پنجم بیان می‌شود. با توجه به مطالب بیان شده در این فصل، در انجام این رساله، بر حصول دقت مناسب، توصیه منابع جدید، گوناگونی لیست پیشنهادی در توسعه سیستم توصیه‌گر تاکید می‌شود. برای ارزیابی مدل پیشنهادی جهت بکارگرفته شده در سیستم توصیه‌گر از معیار مفید بودن منابع پیشنهادی و یا به عبارتی صحت پیشنهادها (precision, recall, f1-measure) استفاده می‌شود.

# فصل سوم

# روش پیشنهادی مدلسازی رفتار کاربر برای توسعه یک سیستم توصیه‌گر

## ۳-۱- مقدمه

نحوه مدلسازی پروفایل کاربران و نیز الگوریتم بکارگیری آن دو مسئله اساسی برای توسعه یک سیستم توصیه‌گر می‌باشند. ایجاد مدل مناسب از پروفایل کاربر، تخمین مناسبی از رفتار کاربر در انتخاب منابع را بدست می‌دهد. بدین ترتیب مدلسازی پروفایل کاربر نقش اساسی در کارائی سیستم-های توصیه‌گر دارد. در این رساله نیز به مدلسازی پروفایل کاربر پرداخته می‌شود.

همانطور که در فصل اول بیان شد، مدلسازی پروفایل کاربر به دو دسته عمده، **مدلسازی ساختاری و مدلسازی رفتاری** تقسیم می‌شوند. مدلسازی ساختاری صرفا منعکس کننده مولفه‌ها و ویژگی‌های داده‌ای موجود در تراکنش انجام شده توسط کاربر می‌باشد و میزان موفقیت سیستم توصیه‌گر به مکانیسم تولید پیشنهاد وابسته می‌باشد. اما در مدلسازی رفتاری، غالبا با استفاده از چارچوب شبکه بیزین [۲۹]، مولفه‌ها و ویژگی‌های موثر بر رفتار کاربر در انجام تعامل با سیستم، شناسایی و استنتاج می‌شوند. بدین ترتیب این رویکرد مدلسازی، ضمن در اختیار گرفتن مولفه‌های موجود در پروفایل کاربر، مکانیسم تولید پیشنهاد را نیز ایجاد می‌کند.

در رویکرد مدلسازی رفتاری، سلیقه کاربر به عنوان یک مولفه تاثیر گذار بر رفتار کاربر در مدل گرافیکی احتمالاتی تعیین می‌شود. سپس با استنتاج پارامترهای مدل، سلایق کاربر شناسایی و جهت تولید پیشنهاد منابع، بکار گرفته می‌شود. مسئله اساسی که در روش‌های موجود به آن توجه نشده-است، میزان علاقه (تمایل) کاربر به هر یک از سلایق می‌باشد. به عبارتی در روش‌های موجود به استنتاج سلایق کاربر پرداخته شده‌است و اولویت (اهمیت) هر یک از سلایق از نظر کاربر نادیده گرفته شده‌است و تاثیر سلایق بر رفتار کاربر ( انتخاب آینده کاربر) یکسان می‌باشد.

با توجه به بخش ۱-۳، در این رساله برای ایجاد مدل رفتاری، از زوج مولفه <**سلیقه ، تمایل**> می‌شود. مولفه سلیقه نشان می‌دهد که کاربر چه گروه‌هایی از منابع را علاقه‌مند می‌باشد و مولفه تمایل میزان اهمیت هر یک از سلایق در یک زمان معین را نشان می‌دهد. در روش پیشنهادی، پروفایل کاربر در یک چارچوب شبکه بیزین به نام فرآیند مدلسازی دریکله بکار گرفته می‌شود. با

استفاده از فرآیند مدلسازی دریکله [۴۱]، ضمن استنتاج سلایق کاربر، اهمیت هر یک از آنها نیز محاسبه می‌شود. فرآیند مدلسازی دریکله شامل دو مولفه فرآیند خوشه‌بندی افزایشی و توزیع دریکله می‌باشد. فرآیند خوشه‌بندی افزایشی - که به فرآیند رستوران چینی معروف می‌باشد - برای استخراج سلایق کاربر مورد استفاده قرار می‌گیرد. این فرآیند امکان پویایی سلیقه‌ها در طول زمان را فراهم می‌سازد. توزیع دریکله نیز برای محاسبه اهمیت هر یک از سلیقه‌ها (تمایل کاربر) مورد استفاده قرار می‌گیرد. این توزیع نیز امکان پویایی تمایل کاربر در طول زمان را فراهم می‌سازد. در روش پیشنهادی برای برای مدلسازی کاربر، با استفاده از فرآیند مدلسازی دریکله این دو مولفه در چارچوب شبکه بیزین در کنار یکدیگر قرار داده می‌شوند.

برای ایجاد یک سیستم توصیه‌گر بر مبنای مدل پیشنهادی، ابتدا با بکارگیری پروفایل کاربران در فرآیند مدلسازی دریکله، مدل رفتاری (بردار <**سلیقه ، تمایل**>) آنها ایجاد می‌شود. سپس با استفاده از معیار شباهت مبتنی بر بردار <**سلیقه ، تمایل**>، کاربران مشابه استخراج می‌شوند. در پایان، منابع کاندید براساس اولویت‌دهی کاربران مشابه مرتب و به کاربر متقاضی پیشنهاد می‌شوند.

در ادامه این فصل، ابتدا در بخش ۳-۲، مطالعاتی در مورد ضرورت در نظر گرفتن اهمیت سلایق کاربر انجام شده‌است. در بخش۳-۳، علائم عمومی بکار گرفته شده در بیان ریاضی مدل پیشنهادی آمده‌است. در بخش ۳-۴ نیز چارچوب شبکه بیزین بکار گرفته شده جهت ایجاد مدل رفتاری از پروفایل کاربر، معرفی شده‌است. روش پیشنهادی برای مدلسازی کاربر نیز در بخش ۳-۵ بیان می‌شود. در پایان، بکارگیری مدل پیشنهادی در توسعه یک سیستم توصیه‌گر آمده‌است.

## ۳-۲- تحلیل مولفه‌های موثر بر رفتار کاربر در انتخاب منابع (نوآوری)

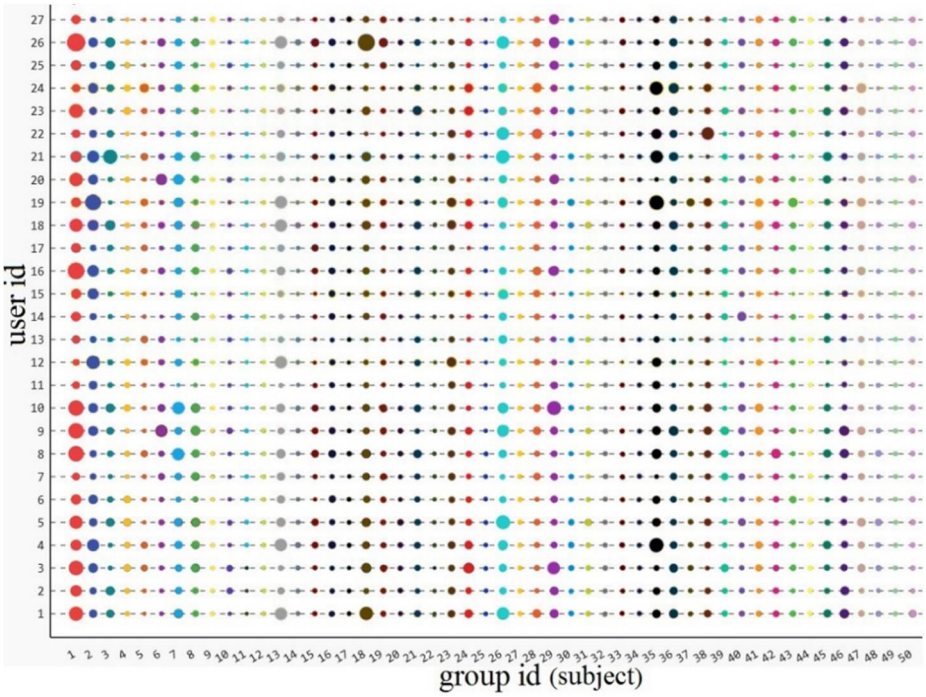
عوامل موثر بر رفتار کاربر در انتخاب منابع ( در سیستم‌های توصیه‌گر ) را می‌توان به دو دسته مولفه‌های مشاهده‌ای[1] و مولفه‌های پنهان[2] تقسیم نمود. مولفه‌های مشاهده‌ای شامل پارامترهای قابل

[1] - Observations
[2] - Latent

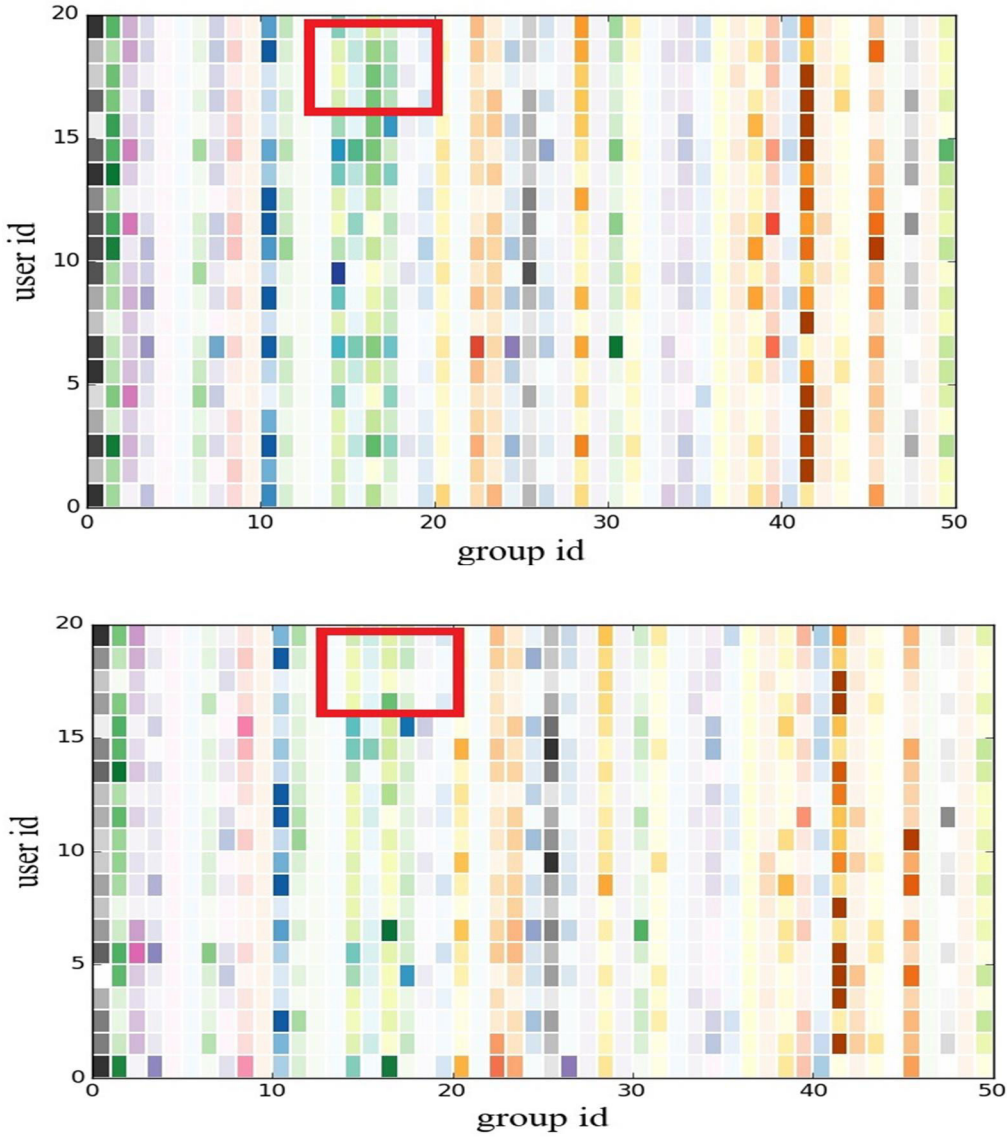
اندازه‌گیری از بستر تعامل می‌باشند. این مولفه غالبا شامل ویژگی کاربر، ویژگی منابع، زمان تعامل، مکان تعامل، و بازخورد می‌باشند. این مولفه‌ها به عنوان اجزاء تشکیل دهنده پروفایل کاربر گردآوری می‌شوند. مولفه‌های پنهان شامل پارامترهای استنتاجی از پروفایل کاربر می‌باشند. این پارامترهای غالبا براساس مولفه‌های مشاهده‌ای محاسبه می‌شوند. در این رساله، سلیقه و علاقه (تمایل) کاربر به عنوان دو مولفه اساسی (از گروه مولفه‌های پنهان) جهت مدلسازی رفتار کاربر در انتخاب منابع، مورد استفاده قرار می‌گیرند. مولفه سلیقه بیان کننده گروه‌هایی از منابع مشابه می‌باشد که کاربر غالبا در طول زمان از آن گروه‌ها منابعی را انتخاب می‌کند. مولفه علاقه (تمایل) نیز بیان کننده میزان اهمیت هر یک از سلیقه‌های کاربر می‌باشد. با توجه به فرضیه رساله، با در نظر گرفتن زوج مولفه‌های پنهان <**سلیقه ، تمایل**> (که از پروفایل کاربران استخراج و محاسبه می‌شوند) بدنبال ایجاد مدل مناسب از رفتار کاربر در انتخاب منابع می‌باشیم.

برای تحلیل رفتار کاربران در انتخاب منابع و ضرورت در نظر گرفتن دو مولفه سلیقه و تمایل، آزمایشاتی انجام شده‌است که در ادامه نتایج آنها بیان می‌شود. برای انجام آزمایش از پایگاه داده NewsTweet استفاده شده‌است. در این پایگاه داده (که شرح کامل آن در بخش ۴-۱ آمده‌است) شامل توئیت‌های خبری مطالعه شده توسط کاربران مختلف می‌باشد. مولفه‌های مشاهده‌ای شامل شناسه کاربر، شناسه توئیت، زمان بازدید، و متن توئیت می‌باشد. برای استنتاج مولفه سلیقه، با استفاده از مدل LDA [۹۳] موضوعات موجود در مجموعه مستندات توئیت‌ها استخراج می‌شود. سپس توئیت‌ها براساس موضوعات مشترک دسته‌بندی می‌شوند. در این آزمایش، موضوعات (که هر یک بیان‌کننده مجموعه‌ای از توئیت‌های مشابه است)، به عنوان سلایق کاربران در نظر گرفته می‌شوند. نتیجه آزمایش در شکل ۳-۱ آمده‌است. در این شکل محور افقی، شناسه سلایق (موضوعات) و محور عمودی، شناسه کاربران را نشان می‌دهند. هر دایره توپر انتخاب منبع توسط کاربر از سلیقه نظیر آن می‌باشد. اندازه هر دایره نیز متناسب با تعداد انتخاب‌ها از هر سلیقه می‌باشد. با توجه به این شکل، تعداد منابعی که کاربر از هر سلیقه انتخاب می‌کند، یکسان نمی‌باشد. در ادامه این آزمایش، از

دو پارامتر تعداد منابع انتخاب شده از هر سلیقه و زمان انتخاب هر منبع برای محاسبه میزان علاقه کاربر به هر سلیقه استفاده شده‌است. نتیجه این آزمایش در شکل ۳-۲ آمده‌است. با توجه به این شکل، میزان غلظت رنگ هر بلوک، میزان علاقه کاربر به سلیقه نظیر را نشان می‌دهد. همچنین، این شکل شامل نتیجه اجرای دو بازه زمانی متوالی می‌باشد که قسمت مشخص شده از نمودارها بیان-کننده تغییرات علاقه کاربر در طول زمان است. این دو آزمایش نشان می‌دهد که رفتار کاربر در انتخاب منابع متاثر از دو مولفه از زوج مولفه <**سلیقه ، تمایل**> می‌باشد.




شکل ۳-۱: نتیجه آزمایش مربوط به تاثیر مولفه‌های سلیقه و تمایل در انتخاب منابع را نشان می‌دهد. محور افقی، بیان کننده شناسه سلیقه کاربر می‌باشد.




شکل ۳-۲: نمایش آزمایش مربوط به میزان تمایل کاربر به انتخاب منابع از هر سلیقه در دو بازه زمانی متوالی. محور افقی، بیان کننده شناسه سلیقه کاربر می‌باشد.

برای بیان مدل پیشنهادی ابتدا مفاهیم منبع، فعالیت، سلایق کاربر، و علایق کاربر تعریف می‌شوند. همچنین نمادهای استفاده شده جهت بیان مدل پیشنهادی در جدول ۳-۱ آمده‌است.

یک منبع با یک بردار ویژگی شامل کلمات معرفی می‌شود. به عنوان مثال برای پایگاه داده NewsTweet، یک منبع، از کلمات کلیدی موجود در متن توئیت‌ها تشکیل می‌شود. البته عملیات پردازش متن نظیر ریشه کلمه و حرف ربط برای توئیت‌ها انجام می‌شود.

در بیان مدل پیشنهادی، منبع با استفاده از نماد $r$ نشان داده می‌شود. یک فعالیت عبارت است از انتخاب یک منبع $(r)$ در زمان معین $(d)$ توسط یک کاربر $(u)$. فعالیت با نماد $a$ نشان داده می‌شود. فعالیت‌های انجام شده توسط کاربر به گروه‌های مجزا ( هر یک شامل فعالیت‌های مشابه هستند) تقسیم می‌شوند که مجموعه این گروه‌ها به عنوان سلایق کاربر تعریف می‌شوند. سلیقه $i$ام کاربر $u$ با $t_i^u$ نشان داده می‌شود. میزان اهمیت هر یک از سلیقه‌ها (احتمال انتخاب منبع از هر سلیقه) به عنوان علاقه (تمایل) کاربر تعریف می‌شود. بدین ترتیب احتمال انتخاب منبع از سلیقه $i$ام با نماد $\theta_i^u$ نشان داده می‌شود. همچنین مجموعه سلایق کاربر $u$ با نماد $T^u$ و مجموعه علایق کاربر $u$ با نماد $\Theta^u$ نشان داده می‌شوند. بدین ترتیب، با فرض اینکه $U = \{u_1, \dots, u_{|U|}\}$ مجموعه کاربران، $R = \{r_1, \dots, r_{|R|}\}$ مجموعه منابع و $A = \{a_1, \dots, a_{|A|}\}$ مجموعه فعالیت‌های کاربران باشند، هدف ما استنتاج و محاسبه دو مجموعه $T^u$ و $\Theta^u$ برای مدلسازی هر کاربر $u$ می باشد.

جدول ۳-۱: نمادهای استفاده شده برای بیان مدل پیشنهادی.

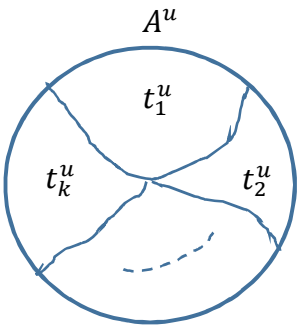
| نماد | شرح |
|---|---|
| $R, r$ | $R$ مجموعه منابع را نشان می‌دهد و $r$ یک منبع خاص می‌باشد |
| $u$ | یک کاربر را نشان می‌دهد |
| $d$ | زمان انجام یک فعالیت را نشان می‌دهد |
| $A^u$ | مجموعه فعالیت‌ها را نشان می‌دهد |
| $t_i, \boldsymbol{t}_j$ | $t_i$ و $\boldsymbol{t}_j$ به ترتیب، سلیقه jام و سلیقه منبع iام را نشان می‌دهند |
| $T^u$ | مجموعه سلایق کاربر $u$ را نشان می‌دهد |
| $\mathcal{F}(.)$ | تابع زمانی (سن) فعالیت را نشان می‌دهد |
| $\mathcal{D}(.)$ | ماتریس شباهت بین فعالیت‌ها را نشان می‌دهد |
| $\Theta^u$ | مجموعه علایق کاربر $u$ را نشان می‌دهد |
| $\beta$ | پارامتر توزیع دریکله برای کنترل توزیع علایق کاربر $u$ |
| $\alpha$ | پارامتر توزیع سلایق کاربر $u$ |
| $G_0$ | تابع احتمال بکارگرفته شده در فرآیند دریکله |
| $\Psi$ | مجموعه‌ای از پارامترها را نشان می‌دهد |

## ۳-۳- فرآیند مدلسازی دریکله

خوشه‌بندی به عنوان یکی تکنیک‌های معروف در توسعه سیستم‌های توصیه‌گر مورد استفاده قرار می‌گیرد. به عنوان مثال از این تکنیک برای خوشه‌بندی کاربران و یا منابع در رویکرد پالایش مشارکتی استفاده می‌شود. هدف این کاربرد، محدود کردن فضای جستجو جهت تعیین همسایگی (چیزهای مشابه) می‌باشد. دو مسئله اساسی در این تکنیک، تعیین تعداد خوشه‌ها و امکان پویایی خوشه‌ها می‌باشد [۴۴, ۹۹]. از اینرو استفاده از چارچوب مدل‌های بیزین ناپارامتریک جهت انجام خوشه‌بندی، به عنوان یکی از راه حل‌های مفید برای غلبه بر این مشکلات می‌باشد. مدل بیزین ناپارامتریک، براساس مشاهدات (پایگاه داده موجود)، تعدادی پارامتر پنهان تعریف می‌کند که هر یک از این پارامترها نماینده یک خوشه از داده‌ها می‌باشند. این پارامترها با مشاهدات جدید در طول زمان پویا می‌باشند و از روی داده‌ها استنتاج (posterior inference) می‌شوند [۴۱].

مدل ترکیبی فرآیند دریکله (DPMM)[1] یکی از توسعه‌های مدل بیزین ناپارامتریک می‌باشد. این فرآیند مدلسازی از تعداد نامحدود پارامتر پنهان (نماینده خوشه‌ها) تشکیل شده‌است [۴۰]. این مدل از دو مولفه اصلی تشکیل شده‌است که در ادامه ضمن تشریح هر یک، مدل DPMM معرفی می‌شود.

**فرآیند دریکله.** فرآیند دریکله توزیع توابع احتمال در یک فضای مسئله را نشان می‌دهد. با توجه به شکل ۳-۳، فضای مسئله $A^u$ که مجموعه داده‌های کاربر $u$ را نشان می‌دهد به افرازهای $\boldsymbol{t_i^u}$ تقسیم




شکل ۳-۳: افرازهای فضای مسئله جهت نمایش توزیع دریکله.

[1] - Dirichlet Process Mixture Model

می‌شود. تابع احتمال $G_0$ نیز سنجه عمومی برای افرازهای ایجاد شده از فضای $A^u$ می‌باشد.

بدین ترتیب فرآیند دریکله توزیع تابع احتمال $G_0$ در هر یک از افرازها را مطابق رابطه ۳-۱ ایجاد می‌کند. پارامتر $\alpha$، توزیع افرازها را کنترل می‌کند. با فرض اینکه $\theta_i^u \sim G_0(t_i^u)$، پارامتر توزیع هر افراز باشد در اینصورت توزیع دریکله را می‌توان بصورت رابطه ۳-۲ نشان داد. در حقیقت فرآیند دریکله، توزیعی روی توابع توزیع پایه $G_0$ را تعریف می‌کند. تابع توزیع گاوسی یا چند جمله‌ای می‌توانند نمونه‌هایی از تابع توزیع پایه $G_0$ باشند [۹۹].

$$\begin{aligned} &G \sim DP(\alpha, G_0)\,, \\ &(G(t_1^u), \ldots, G(t_k^u)) \sim DP(\alpha G_0(t_1^u), \ldots, \alpha G_0(t_k^u)) \end{aligned} \tag{۳-۱}$$

$$\theta \sim \mathrm{Dir}(\alpha) = \prod_{j=1:|\theta|} \theta_j^{\alpha_j} \tag{۳-۲}$$

**فرآیند رستوران چینی**(ddCRP)[1]. فرآیند رستوران چینی [۳۹] یک فرآیند خوشه‌بندی افزایشی مبتنی بر چارچوب بیزین ناپارامتریک را فراهم می‌کند. در این فرآیند خوشه‌بندی ضمن اینکه داده‌ها براساس میزان شباهت در کنار یکدیگر قرار می‌گیرند، به احتمال $\alpha$ می‌توانند خوشه جدید ایجاد نمایند. به عبارتی، در این روش، داده‌های مشابه به یکدیگر متصل شده و یک خوشه را تشکیل می‌دهند. همچنین به ازای مشاهده داده جدید، امکان ادغام خوشه‌های جاری نیز وجود دارد. رابطه ۳-۳، نمای کلی این روش را نشان می‌دهد. در این رابطه، $\mathbf{D}(.)$ ماتریس شباهت بین داده‌ها می‌باشد. با توجه به رابطه ۳-۳، احتمال اینکه داده **i**ام به داده **j**ام متصل شود ( یا به عبارتی در یک خوشه قرار گیرند)، متناسب با شباهت بین این دو داده و ضریب توزیع خوشه‌ها می‌باشد.

$$p(t_i = j \mid D(.), \alpha) = \begin{cases} D(i,j) & if\ i \neq j \\ \alpha & \end{cases} \tag{۳-۳}$$

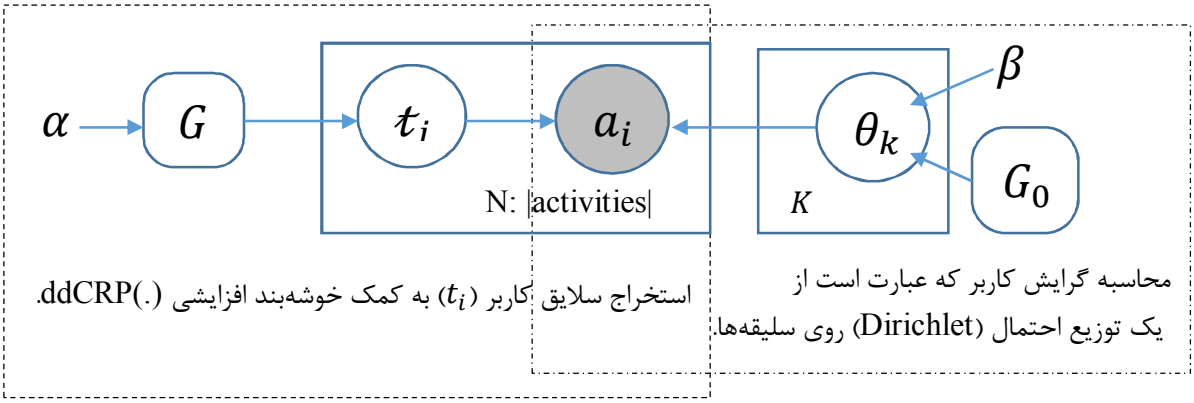
[1] - Distance Dependent Chienece Restaurant Process

بدین ترتیب در این رساله، یک مدل احتمالاتی گرافیکی برای مدلسازی رفتار کاربر پیشنهاد می‌شود. روش پیشنهادی با بکارگیری دو مولفه فرآیند دریکله و فرآیند رستوران چینی در چارچوب DPMM، مدل رفتاری کاربر را ایجاد می‌کند. در بخش بعدی، بیان مدل گرافیکی و استنتاج پارامترها تشریح می‌شود.

## ۳-۴- ارائه مدل پیشنهادی کاربر (نوآوری)

در روش پیشنهادی برای مدلسازی رفتار کاربر، هدف ایجاد بردار <**سلیقه ، تمایل**> برای کاربر می‌باشد. با توجه به جدول ۳-۱، که مولفه سلیقه و تمایل کاربر $u$ را به ترتیب با $\boldsymbol{t_i^u}$ و $\theta_i^u$ نمایش داده می‌شود، برای ایجاد مدل کاربر باید بردار $\prec (\boldsymbol{t_1^u}, \theta_1^u), \dots, (\boldsymbol{t_k^u}, \theta_k^u) \succ$ به ازای هر کاربر $u$ ساخته شود. رفتار کاربر با استفاده از مدل گرافیکی احتمالاتی DPMM مدل می‌شود. در این مدل گرافیکی احتمالاتی، منابع انتخاب شده توسط کاربر $u$ ($A^u$)، مولفه سلیقه ($\boldsymbol{t_i^u}$)، و مولفه تمایل ($\theta_i^u$) به عنوان پارامترهای مدل تعیین می‌شوند. پارامترهای سلیقه و تمایل به عنوان فاکتورهای پنهان در مدل استنتاج می‌شوند. همانطور که قبلا بیان شد، مولفه سلیقه نشان می‌دهد که کاربر چه گروه‌های از منابع را در طول زمان انتخاب نموده‌است. به عبارتی این مولفه، نیازمندی‌های کاربر در طول زمان را نشان می‌دهد. مولفه‌های تمایل نیز، سهم هر یک از سلیقه‌ها در انتخاب منابع (در زمان آینده) توسط کاربر را نشان می‌دهند. به ازای زمان معین، مولفه تمایل میزان اهمیت هر یک از سلیقه‌ها را از دیدگاه کاربر محاسبه می‌کنند. سپس براساس اهمیت هر یک از سلیقه‌ها، تخمینی از انتخاب منابع انجام می‌شود. در این پژوهش، برای محاسبه مولفه تمایل از دو پارامتر تعداد منابع انتخاب شده از هر سلیقه و زمان هر انتخاب استفاده می‌شود. این دو پارامتر میزان و تازگی انتخاب‌های انجام شده از هر سلیقه را نشان می‌دهند. شکل ۳-۴ مدل گرافیکی احتمالاتی مورد استفاده جهت مدلسازی رفتار کاربر در این رساله را نشان می‌دهد. در این شکل بلاک سلیقه، سلایق کاربر را با استفاده از داده‌ها استنتاج می‌کند

و بلاک تمایل نیز با توجه به داده‌ها و سلایق، تمایل کاربر به ازای هر سلیقه را محاسبه می‌کند. شرح پارامترهای این مدل در جدول ۳-۱ آمده‌است.




شکل ۳-۴: مدل گرافیکی احتمالاتی برای نمایش مدل پیشنهادی رفتار کاربر.

**عملکرد مدل.** در این مدل گرافیکی احتمالاتی، سلایق کاربر (مولفه‌های $t_i^u$) از توزیع فرآیند رستوران چینی (ddCRP) پیروی می‌کند و به عنوان فاکتورهای پنهان از روی مشاهدات ($a_i$ ها) استخراج می‌شود که در شکل ۳-۴، به اختصار با π نشان داده شده‌است. به عبارتی، این فاکتورهای پنهان، خوشه‌بندی از مشاهدات (انتخاب‌های کاربر) را انجام می‌دهند. پس از استخراج سلایق کاربر، مولفه تمایل کاربر ($\theta_i^u$) با استفاده از توزیع دریکله به ازای هر خوشه محاسبه می‌شود. بنابراین، عملکرد مدل را می‌توان به ترتیب ذیل بیان نمود:

- به ازای هر فعالیت کاربر $u$، ($a_i^u \in A^u$): سلیقه این فعالیت شناسایی می‌شود ($t_i^u \sim ddCRP(.)$).
- به ازای هر سلیقه کاربر $u$، ($t_k^u \in T^u$): تمایل کاربر به این سلیقه محاسبه می‌شود ($\theta_k^u \sim Dir(.)$).
- به ازای یک فعالیت جدید ($a$):
  - ابتدا سلیقه آن شناسایی می‌شود: $t_a^u \sim ddCRP(.)$.
  - سپس احتمال انجام فعالیت محاسبه می‌شود: $a \sim Cat(\theta_{t_a^u}^u)$

بیان ریاضی عملکرد مدل پیشنهادی در رابطه ۳-۴ آمده‌است. پارامترهای $\alpha$، $\beta$، $G_0$، و $D$ به عنوان پارامترهای اولیه (ابر-پارامتر) مدل بوده ($\psi=\{\alpha, \beta, G_0, D\}$) و مقادیر آنها قبل از فرآیند محاسبه

تعیین می‌شود. پارامتر $D$ که بیان کننده ماتریس شباهت می‌باشد، براساس مقایسه بین منابع انتخاب شده توسط کاربر محاسبه می‌شود. شرح محاسبه پارامتر نیز در بخش ۴-۴-۲ آمده است. پارامترهای $\mathbf{t}_{\mathrm{i}}^{\mathrm{u}}$ و $\theta_i^u$ جزء پارامترهای پنهان مدل بوده و باید محاسبه شوند. با توجه به بیان ریاضی عملکرد مدل پیشنهادی، برای تخمین مقادیر پارامترهای مدل از استنتاج ثانویه (*posteriori inference*) استفاده می‌شود.

$$\begin{aligned} &G \sim ddCRP(\alpha) \\ &\mathbf{t}_{\mathrm{i}}^{\mathrm{u}} \mid G(.) \\ &\theta_{\mathrm{k}}^{\mathrm{u}} \mid \beta, G_0 \sim Dir(\beta, G_0) \\ &a \mid \mathrm{t}_{\mathrm{a}}^{\mathrm{u}} = z, \theta_{\mathrm{z}}^{\mathrm{u}} \sim Cat(\theta_{\mathrm{z}}^{\mathrm{u}}) \end{aligned} \tag{۴-۳}$$

**استنتاج پارامترهای مدل.** برای تخمین پارامترهای پنهان، ابتدا با استفاده از استنتاج ثانویه، استخراج سلیقه کاربر انجام می‌شود و سپس محاسبه تمایل انجام می‌شود. استخراج سلیقه کاربر به بیان ریاضی، عبارت است از تعیین سلیقه (خوشه) هر یک از فعالیت‌های کاربر که با رابطه ۳-۵ بیان می‌شود. در این رابطه، احتمال تعلق فعالیت $a_i^*$ به سلیقه $k$ام با فرض معین بودن مجموعه پارامترهای اولیه ($\psi$) و سلیقه سایر فعالیت‌ها محاسبه می‌شود. از آنجاییکه منبع $a_i^*$ در مخرج کسر از رابطه ۳-۵ ظاهر می‌شود، از اینرو محاسبه این رابطه با تخمین مقدار صورت کسر انجام می‌شود. بدین ترتیب، رابطه ۳-۵ بصورت رابطه ۳-۶ بازنویسی می‌شود.

$$\mathrm{p}(t_i^* = k | t_{-i}, A^u, \psi) = \frac{p(t_i^* = k | t_{-i}, A^u) p(A^u | (t_i^* = k, t_{-i}), \psi)}{\sum_{t'} p(A^u | t', \psi)\, p(t' | \psi)} \tag{۵-۳}$$

$$\begin{aligned} &\mathrm{p}(t_i^* = k | t_{-i}, A^u, \psi) \propto \\ &\mathrm{p}(t_i^* = k | t_{-i}, \psi)\, p(a_i^* | A_{-i}^u, (t_i^* = k, t_{-i}), \psi) p(A_{-i}^u | (t_i^* = k, t_{-i}), \psi) \end{aligned} \tag{۶-۳}$$

با توجه به اینکه عبارت $p(A_{-i}^u | (t_i^* = k, t_{-i}), \psi)$ از رابطه ۳-۶ در فرآیند تعیین تعلق سلیقه فعالیت $a_i^*$ ثابت می‌باشد، بنابراین از آن صرفنظر شده و رابطه ۳-۶ بصورت رابطه ۳-۷ بازنویسی می‌شود. در

رابطه ۳-۷، عبارت $\mathrm{p}(t_i^* = k|t_{-i},\psi)$ و $p(a_i^*|A_{-i}^u,(t_i^* = k,t_{-i}),\psi)$، به ترتیب احتمال قرارگرفتن فعالیت $a_i^*$ در سلیقه $k$ام را محاسبه می‌کند.

$$\mathrm{p}(t_i^* = k|t_{-i},A^u,\psi) \propto \mathrm{p}(t_i^* = k|t_{-i},\psi)\, p(a_i^*|A_{-i}^u,(t_i^* = k,t_{-i}),\psi) \qquad (۳-۷)$$

از آنجاییکه رابطه ۳-۷، عمل خوشه‌بندی فعالیت‌های کاربر را انجام می‌دهد، از اینرو از فرآیند رستوران چینی برای خوشه‌بندی فعالیت‌های کاربر استفاده می‌شود. براساس رابطه ۳-۳ و ۳-۷، استنتاج ثانویه برای خوشه‌بندی فعالیت‌های کاربر با استفاده از رابطه ۳-۸ انجام می‌شود. در رابطه ۳-۸، نماد $\mathcal{L}$ بیان کننده عبارت $p(a_i^*|A_{-i}^u,(t_i^* = k,t_{-i}),\psi)$، از رابطه ۳-۷ می‌باشد.

$$\mathrm{p}(t_i^* = k|t_{-i},A^u,\psi) = \begin{cases} \alpha * \mathcal{L} & \mathrm{if}(t_i^* = new\ interest) \\ D(a_i^*,a_{-i}) * \mathcal{L} & \mathrm{if}(t_i^* = \mathrm{prev}\ interest) \\ D(a_i^*,a_{-i}) * \mathcal{L} & \mathrm{if}(t_i^* \neq \mathrm{prev}\ interest) \end{cases} \qquad (۳-۸)$$

با توجه به سه حالت بیان شده در رابطه ۳-۸، مقدار عبارت $p(a_i^*|A_{-i}^u,(t_i^* = k,t_{-i}),\psi)$، بصورت ذیل محاسبه می‌شود:

- $\mathrm{if}(t_i^* = new\ interest)$: در این حالت فعالیت $a_i^*$ در یک خوشه (سلیقه) جدید قرار می‌گیرد و چون تنها عضو این خوشه همان فعالیت $a_i^*$ می‌باشد بنابراین مقدار $\mathcal{L}$ برابر با یک فرض می‌شود.
- $\mathrm{if}(t_i^* = \mathrm{prev}\ interest)$: در این حالت فعالیت $a_i^*$ به همان خوشه (سلیقه) قبلی تعلق می‌گیرد. در اینصورت نیاز به محاسبه مجدد شانس انتخاب (likelihood) برای این فعالیت نمی‌باشد.
- $\mathrm{if}(t_i^* \neq \mathrm{prev}\ interest)$: در این حالت، فعالیت $a_i^*$ به یکی از خوشه‌های موجود دیگر تعلق می‌گیرد. در اینصورت باید مقدار شانس انتخاب این فعالیت (عبارت $\mathcal{L}$) محاسبه شود. برای محاسبه عبارت $\mathcal{L}$، فرض می‌شود که $\ell$ و $k$ به ترتیب خوشه جاری فعالیت $a_i^*$ و خوشه جدید آن باشند. همچنین تابع $\mathrm{Z(x)}$ مجموعه فعالیت‌هایی را برگشت می‌دهد که متعلق به خوشه x باشد. بدین ترتیب مقدار عبارت $\mathcal{L}$ (شانس انجام فعالیت $a_i^*$) بصورت رابطه ۳-۹، محاسبه می‌شود.

$$p(a_i^*|A_{-i}^u,(t_i^*=k,t_{-i}),\psi)=p\left(a_i^*\middle|Z(\ell)_{-i},Z(k)_{-i},(t_i^*=k,t_{-i}),\psi\right)=$$
$$\frac{p(Z(k,\ell)|.)}{p\left(Z(\ell)_{-i}\middle|.\right)*p(Z(\mathrm{k})_{-i}|.)}, \qquad (۹-۳)$$
$$p(Z(x)|.)=\sum_{k=1}^{K}p(Z(x)|\theta_k)*p(\theta_k|\psi),$$

بدین ترتیب با اعمال روابط ۳-۸ و ۳-۹ بر مجموعه داده‌ها، تعیین مولفه سلیقه کاربر انجام می-شود. به عبارتی با بکارگیری مجموعه داده‌ها در ساختار گرافیکی احتمالاتی و انجام خوشه‌بندی، سلایق هر کاربر - که بیان کننده مجموعه‌ای از منابع مشترک انتخاب شده توسط کاربر است- استخراج می‌شود.

**مولفه تمایل.** همزمان با استخراج مولفه سلیقه، مولفه تمایل کاربر نیز باید محاسبه گردد. به عبارتی به ازای استنتاج هر پارامتر $\boldsymbol{t}_k^u$، باید مولفه تمایل یا همان پارامتر $\theta_k^u$ نظیر آن نیز محاسبه شود. از آنجاییکه مولفه تمایل بیان کننده سهم (تاثیر) هر سلیقه در انتخاب کاربر می‌باشد، از اینرو برای محاسبه آن از یک تابع احتمال استفاده می‌شود. این تابع احتمال اولویت نسبی هر سلیقه نسبت به سایر سلیقه‌های کاربر را محاسبه می‌کند. برای بیان ریاضی این تابع، ابتدا اولویت هر سلیقه براساس تعداد و زمان فعالیت‌های انجام در آن سلیقه، محاسبه می‌شود. رابطه ۳-۱۰، نحوه محاسبه اولویت هر سلیقه را نشان می‌دهد. در این رابطه تابع $Weight(.)$ با استفاده از برچسب زمانی (*timestamp*) هر منبع انتخاب شده (*activity*)، ارزش وزنی آنرا محاسبه می‌کند. تابع $G_0$ به عنوان تابع پایه‌ای (تابع احتمال) توزیع سلایق کاربر مورد استفاده قرار می‌گیرد. این تابع سهم هر سلیقه را تعیین می‌کند.

$$G_0(interest)=\frac{\sum_{activity\,\epsilon\,interest}Weight(activity)}{\sum_{j=1}^{|T|}\left(\sum_{activity\,\in\,interest_j}Weight(activity)\right)} \qquad (۱۰-۳)$$
$$Weight(activity)=exp(-activity.timestamp)$$

با توجه به رابطه ۳-۱۰، برای بکارگیری توابع احتمال و تعیین توزیع هر یک از سلایق، از تابع توزیع دریکله استفاده می‌شود. توزیع دریکله دارای دو مولفه تابع احتمال ($G_0$) و پارامتر وزن‌دهی$\beta$ می‌باشد. رابطه ۳-۱۰ به عنوان تابع احتمال این توزیع بکار می‌رود و مقدار پارامتر $\beta$ بصورت پیش

فرض برای همه یکسان فرض می‌شود. رابطه ۳-۱۱، توزیع دریکله مورد استفاده برای تعیین توزیع سلایق کاربر را نشان می‌دهد.

$$(\theta_1, \dots, \theta_k) = Dir(\beta, G_0) \qquad (۱۱-۳)$$

**ایجاد مدل.** با توجه به روابط بیان شده (روابط ۳-۸، ۳-۹، و ۳-۱۱)، ایجاد مدل کاربر عبارت است از تخمین پارامترهای (پنهان) $\boldsymbol{t}_k^u$ و $\theta_k^u$ که با اعمال این روابط بر روی داده‌ها محاسبه می‌شوند. الگوریتم بیان شده در شکل ۳-۵، فرآیند ایجاد مدل را نشان می‌دهد. این الگوریتم به *Gibbs Sampling* معروف می باشد که با اعمال روابط ۳-۸، ۳-۹، و ۳-۱۱ بر روی داده‌ها (منابع انتخاب شده توسط کاربران) در یک ساختار تکرار، استنتاج پارامترهای (پنهان) $\boldsymbol{t}_k^u$ و $\theta_k^u$ را انجام می‌دهد.

**Algorithm 1: Gibbs sampling algorithm for approximating parameters**

**Input**: $A^u$, Ψ = {α, β, D(.), G0}.
**Output**: model parameter Θ as user preferences distribution and latent clusters (tables) T with their members.
**Initialization**:
  N= size($A^u$).
  Initialize the latent clusters (tables) T.
  Randomly place activities on initial tables.
**While** not converged do
  **For** i = 1 to N do
    Select ith activity and ignore its table assignment.
    dist_vector = {}
    **For** k = 1 to | T | do
      $p_{ik}$ = Compute the probability of assigning the ith activity to the $k^{th}$ table with Eq. 3-8.
      dist_vector = dist_vector ∪ $p_{ik}$
    **EndFor**
    $t_i^*$= Choose the destination table of the ith activity according to dist vector.
    **If** $t_i^*$ is a new table then
      T = T ∪ $\{t_i^*\}$
    **EndIf**
    Compute model parameter θ using Eq. 3-11.
    Update hyper-parameter α.
  **EndFor**
**EndWhile**

شکل ۳-۵: فرآیند محاسبه پارامترها و ایجاد مدل.

## ۳-۵- بکارگیری مدل پیشنهادی جهت ارائه توصیه (نوآوری)

با توجه به فرآیند بیان شده در شکل ۳-۵، پارامترهای مورد نیاز برای ایجاد مدل کاربر محاسبه می‌شوند. مدل کاربر شامل دو پارامتر اساسی سلیقه و تمایل کاربر می‌باشد. این مدل را می‌توان بصورت رابطه ۳-۱۲ نشان داد. در ادامه با بکارگیری رابطه ۳-۱۲، ابتدا با رویکرد محتوایی و سپس با رویکرد ترکیبی، توسعه سیستم توصیه‌گر انجام می‌شود.

$$pref^{u_i} = \{< \boldsymbol{t}_1^{u_i}, \theta_1^{u_i} >, \dots, < \boldsymbol{t}_{k_i}^{u_i}, \theta_{k_i}^{u_i} >\} \qquad (۳-۱۲)$$

### ۳-۵-۱- ایجاد سیستم توصیه‌گر بر اساس مولفه‌های سلیقه و تمایل (نوآوری)

در این مرحله با در دست داشتن مدل کاربر، استفاده از آن در ایجاد یک الگوریتم توصیه بیان می‌شود. این الگوریتم الویت‌دهی منابع کاندید جهت پیشنهاد به کاربر را انجام می‌دهد. برای بکارگیری مدل کاربر و فراهم نمودن توصیه‌ها، مراحل ذیل انجام می‌شود:

- با استفاده از فرآیند خوشه‌بندی رستوران چینی (ddCRP) هر یک از منابع کاندید دسته‌بندی می‌شوند. رابطه ۳-۸ و ۳-۹ بر مجموعه داده‌های کاندید اعمال می‌شوند و سلیقه هر یک از منابع تعیین می‌شود. به عبارتی احتمال تعلق هر منبع به سلیقه کاربر ($\boldsymbol{t}_k^u$) تعیین می‌شود.
- پس از تعیین سلیقه هرمنبع کاندید، از مولفه اولویت ($\theta_k^u$) هر سلیقه برای تعیین میزان محبوبیت منبع مورد نظر استفاده می‌شود.

این دو مرحله با استفاده رابطه ۳-۱۳ فرموله می‌شوند. در این رابطه میزان محبوبیت هر منبع ($r$) از نظر کاربر متقاضی ارزیابی می‌شود. در این رابطه، عبارت‌های بعد از زیکما به ترتیب از سمت چپ عبارتنداز تعیین سلیقه منبع کاندید، اولویت سلیقه نظیر، و شانس انتخاب منبع کاندید. فرآیند بکارگیری مدل کاربر و تهیه لیست پیشنهادی در شکل ۳-۶ آمده‌است.

$$p(r|A^u,\psi) = \frac{1}{p(A^u|\psi)}\sum_{k=1}^{K} p(t_r = k|\psi) \times p(\theta_k|\psi) \times p(r, A^u|\theta_k, \psi) \qquad (۱۳-۳)$$

**Algorithm 2: Prediction process to provide the recommendations.**

**Input**: parameters that specify user model, such as latent clusters (T), preference distribution (Θ) and the resource candidate list.
**Output**: An ordered list for recommendation.
**Initialization**:
N = size(candidate list).
Priority_list = {}
**For** j = 1 to N do
rj = Select jth resource from candidate list.
tj = Choose the destination table of resource rj using Eq. 8.
pj = Calculate the likelihood of activity rj given destination table tj using Eq. 16.
Priority list = priority_list ∪ {rj, pj}
**EndFor**
**Return** Descending sort of priority list.

شکل ۳-۶: فرآیند بکارگیری مدل کاربر و تهیه لیست پیشنهادی.

### ۳-۵-۲- ایجاد سیستم توصیه‌گر ترکیبی (نوآوری)

در بخش قبل با استفاده از مدل پیشنهادی کاربر، یک سیستم توصیه‌گر محتوایی ایجاد شده‌است. در این بخش بکارگیری مدل پیشنهادی برای ایجاد الگوریتم ارائه توصیه ترکیبی با مولفه‌های محتوایی و مشارکتی انجام می‌شود. . با توجه به تحقیقات انجام شده [۵۰, ۵۱] در زمینه بهبود سیستم‌های توصیه‌گر مبتنی بر پالایش مشارکتی، یکی از مسائل اساسی در بهبود آنها، بکارگیری معیار شباهت مناسب در تعیین همسایگی در این دسته از الگوریتم‌ها می‌باشد. در روش پیشنهادی با بکارگیری مدل ایجاد شده برای کاربر، معیار جدیدی برای محاسبه شباهت کاربران و تعیین همسایگی مطرح شده‌است. ایده اصلی روش پیشنهادی بر این مبنا می‌باشد که برخلاف روش‌های موجود [۵۱]، به جای استفاده از ویژگی‌های سطح پایین (امتیازها) از ویژگی‌های سطح بالا (سلیقه کاربران ) برای

محاسبه شباهت استفاده شود. همچنین مولفه تمایل کاربر نیز در تعیین میزان شباهت کاربران، محاسبه می شود.

فرآیند ایجاد سیستم توصیه‌گر ترکیبی پیشنهادی به شرح ذیل می‌باشد:

- بکارگیری پروفایل هر کاربر و ایجاد مدل رفتاری کاربر براساس بخش ۳-۴. به عبارتی با بکارگیری پروفایل هرکاربر مدل رفتاری کاربر مطابق رابطه ۳-۱۲ ایجاد می‌شود.
- با توجه به مدل کاربر بیان شده در رابطه ۳-۱۲، شباهت بین کاربران براساس میزان شباهت بین سلیقه‌ها و اولویت آنها محاسبه می‌شود. برای تعیین شباهت بین کاربران از رابطه ۳-۱۴ استفاده می‌شود. در این رابطه تابع $Similarity(.)$ شباهت دو کاربر $u_x$ و $u_y$ را محاسبه می‌کند. این تابع از توابع $interestSim(.)$ و $preferenceSim(.)$ به ترتیب برای محاسبه میزان شباهت بین سلیقه دو کاربر و میزان اولویت سلایق نظیر استفاده می‌کند.

$$Similarity(u_x, u_y) = Similarity(pref^{u_x}, pref^{u_y}) = \sum_{i=1}^{k_{u_x}} \sum_{j=1}^{k_{u_y}} interestSim\left(\boldsymbol{t}_i^{u_x}, \boldsymbol{t}_j^{u_y}\right) * preferenceSim\left(\theta_i^{u_x}, \theta_i^{u_y}\right) \qquad (۱۴-۳)$$

$$preferenceSim\left(\theta_i^{u_x}, \theta_i^{u_y}\right) = e^{-|\theta_i^{u_x} - \theta_i^{u_y}|}$$

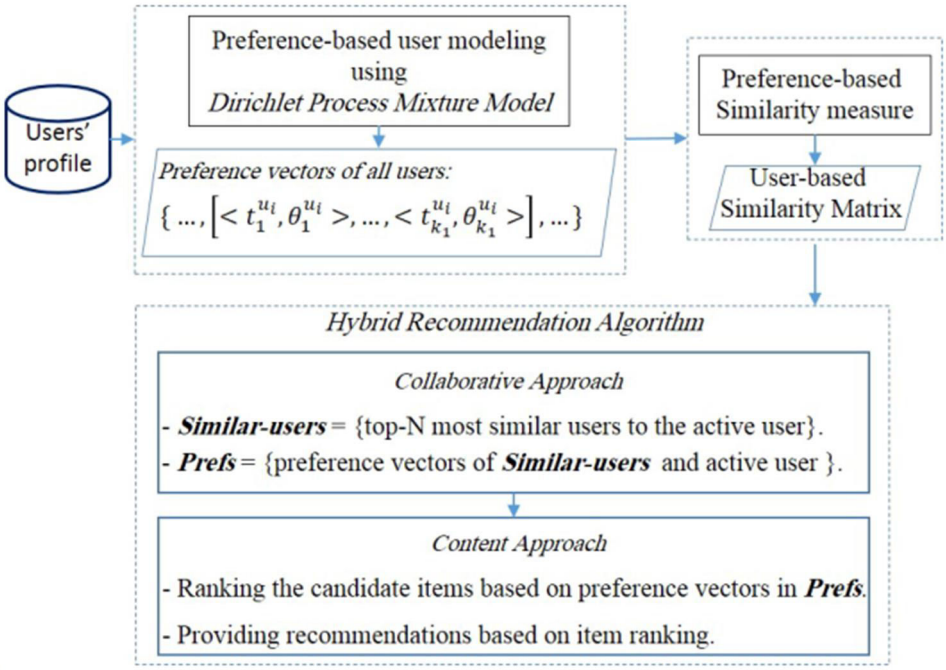
- پس از تعیین کاربران مشابه، با بکارگیری مدل رفتاری کاربران مشابه در رابطه ۳-۱۳، منابع مرتب و به کاربر متقاضی پیشنهاد می‌شوند.

با توجه به مراحل بیان شده، الگوریتم توصیه با رویکرد ترکیبی در شکل ۳-۷ آمده‌است. همچنین شکل ۳-۸ معماری سیستم توصیه‌گر ترکیبی را نشان می‌دهد.

**Algorithm 3: The recommendation algorithm.**

**Input:** preference vector, similarity matrix, active user u, and candidate items.
**Output**: an ordered list for recommendation.
**Initialization:**
Similar_users = select top-N most similar user to u.
Prefs = collect preference vectors of Similar_users and u.
Priority_list = {}
**For** each item in candidate_items
p= Calculate the likelihood of item using Eq. 3-13.
Priority_list = Priority_list ∪ {item, p}.
**EndFor**
**Return** Priority_list in descending order of likelihood.

شکل۳-۷: الگوریتم ارائه توصیه با رویکرد ترکیبی.




شکل ۳-۸: معماری سیستم توصیه‌گر ترکیبی پیشنهادی.

## ۳-۶- جمع‌بندی

در این بخش، فرآیند پیشنهادی برای مدلسازی رفتار کاربر در انتخاب منابع تشریح شده‌است. برای بیان مدل پیشنهادی، ابتدا تحلیلی از رفتار کاربر در انتخاب منابع انجام شده‌است. برای بررسی رفتار

کاربر، از پایگاه داده‌های NewsTweets (که شرح کامل آنها در بخش ۴-۱ آمده‌است) استفاده شده-است. تحلیل رفتار کاربران در انتخاب منابع نشان می‌دهد که گروه‌های مشخصی از منابع از نظر کاربر محبوبیت بیشتری دارند. به عبارتی، این گروه‌ها، نیاز اساسی کاربر را منعکس می‌کنند. از طرفی، هر یک از گروه‌ها در طول زمان، دارای اولویت مشخصی از نظر کاربر می‌باشند. بدین ترتیب برای مدلسازی رفتار کاربر از زوج مولفه <**سلیقه ، تمایل**> استفاده شده‌است.

برای ایجاد مدل کاربر، از فرآیند مدلسازی دریکله استفاده شده‌است. این فرآیند مدلسازی دارای دو مولفه اساسی خوشه‌بند افزایشی و اولویت بندی خوشه‌ها تشکیل شده‌است. مولفه خوشه‌بند افزایشی برای استخراج سلیقه‌های کاربران در طول زمان استفاده می‌شود. مولفه اولویت بندی نیز، اهمیت هر یک از سلیقه‌ها را محاسبه می‌کند. مدل پیشنهادی ابتدا در سیستم توصیه‌گر محتوایی بکار گرفته شد. سپس سیستم توصیه‌گر ترکیبی با مولفه‌های محتوایی و مشارکتی برای بکارگیری مدل پیشنهادی ایجاد شده‌است. در فصل بعدی، برای بررسی عملکرد مدل پیشنهادی در بهبود عملکرد سیستم‌های توصیه‌گر، آزمایشات لازم انجام و نتایج گردآوری شده‌است.

# فصل چهارم

# آزمایش‌ها و تحلیل نتایج

## ۴-۱- مقدمه

در این فصل، آزمایشات مختلف جهت بررسی عملکرد مدل پیشنهادی تشریح می‌شوند. در انجام آزمایشات، ابعاد مختلف مدل پیشنهادی و پارامترهای موثر در عملکرد مدل پیشنهادی مورد بررسی قرار می‌گیرند. در ادامه ابتدا پایگاه داده استفاده شده جهت انجام آزمایشات و معیارهای ارزیابی معرفی می‌شوند. همچنین روش‌های مورد مقایسه‌ای نیز جهت ارزیابی میزان کارائی روش پیشنهادی و انجام مقایسه بیان می‌شوند. سپس آزمایشات مختلف و تحلیل نتایج ارائه می‌شوند.

## ۴-۲- پایگاه داده و محیط پیاده‌سازی

برای ارزیابی روش پیشنهادی از دو پایگاه داده واقعی استفاده شده‌است. پایگاه داده اول، *MovieLens*[1] می‌باشد که در اکثر روش‌های موجود بکارگرفته شده‌است. هر رکورد از این پایگاه داده بیان کننده امتیاز یک کاربر به یک منبع خاص در یک زمان معین می‌باشد. رکوردهای این پایگاه داده را می‌توان بصورت {شناسه کاربر، شناسه فیلم، خلاصه داستان فیلم، امتیاز کاربر، برچسب زمان} نشان داد. پایگاه داده دوم، *NewsTweets*[2] می‌باشد که شامل توئیت‌های مطالعه شده توسط کاربران در طول زمان می‌باشد. این پایگاه داده در بازه می (*May*) ۲۰۱۷ تا آگوست ۲۰۱۷ گردآوری شده-است. رکوردهای این پایگاه داده را نیز می‌توان بصورت {شناسه کاربر، شناسه توئیت، متن توئیت، برچسب زمانی} نشان داد. اطلاعات آماری این دو پایگاه داده در جدول ۴-۱ آمده‌است.

جدول۴-۱: مشخصات آماری پایگاه داده‌های استفاده شده در آزمایشات.

| | **Datasets** | |
|---|---|---|
| **Properties** | *MovieLens* | *NewsTweet* |
| *Number of users* | 4743 | 2018 |
| *Number of resources* | 8787 | 46936 |
| *Number of feedbacks* | 1M | 376085 |

[1] - https://grouplens.org/datasets/movielens
[2] - http://ipdm.shahroodut.ac.ir/downloads/link

برای پیاده‌سازی روش پیشنهادی، از چارچوب برنامه نویسی چندپردازه‌ای استفاده می‌شود. این پیاده‌سازی با *Python* انجام می‌شود. سیستم سخت افزاری بکارگرفته شده دارای پردازنده *Intel Core i28* و حافظه *16GB* می‌باشد.

## ۴-۳- معیارهای ارزیابی و روش‌های مورد مقایسه‌ای

در روش پیشنهادی، به ازای هر بار انجام فرآیند توصیه، تعداد $N$ منبع به ترتیب اولویت به کاربر متقاضی پیشنهاد می‌شوند. برای ارزیابی این لیست پیشنهادی از دو معیار استفاده می‌شود: معیار صحت توصیه‌ها و ترتیب توصیه‌ها. برای بررسی میزان صحت توصیه‌ها پارامترهای دقت و بازخوانی( precision, recall, F1 – measure) اندازه‌گیری می‌شوند. برای بررسی میزان درستی ترتیب منابع پیشنهادی نیز از معیار *DCG*[1] استفاده می‌شود [۴۲, ۴۳, ۹۷, ۹۸].

برای مقایسه نتایج و ارزیابی میزان کارائی روش پیشنهادی، تعدادی از روش‌های معروف انتخاب شده‌اند. این روش‌ها از رویکردهای مشارکتی مبتنی بر شباهت منابع (*item-item based CF*) [۱]، رویکردهای مشارکتی مبتنی بر معیار پنهان (*SVD, NMF, timeSVD, FunkSVD*) [۳۳, ۳۷, ۴۷, ۱۰۰]، و رویکرد محتوایی (*content based filtering*) [۲۲] در انجام توصیه استفاده می‌کنند. این روش‌ها در قالب ابزارهایی نظیر *Librec*[2]، *Surprise*[3]، و *Lenskit*[4] پیاده‌سازی شده‌اند. همچنین الگوریتم ارائه شده در مقاله [۷۴, ۱۰۱] (که *LOGO* نامیده می‌شود) به عنوان یک سیستم توصیه خبر جهت مقایسه انتخاب و پیاده‌سازی شده‌است.

## ۴-۴- آزمایش‌ها و نتایج

در این بخش ضمن بیان ابعاد مختلف ارزیابی روش پیشنهادی، نتایج حاصل از ارزیابی‌ها نظیر صحت و ترتیب توصیه‌ها نیز تشریح می‌شود. همچنین پارامتر کنترل توزیع سلایق کاربر و تاثیر مولفه **تمایل** بر بهبود معیار شباهت در رویکرد مشارکتی نیز بررسی می‌شود.

---

[1] - Discounted Cumulative Gain (DCG)
[2] - https://www.librec.net/
[3] - http://surprise.readthedocs.io/en/stable
[4] - http://lenskit.grouplens.org/

### ۴-۴-۱- ارزیابی صحت مدل پیشنهادی

برای ارزیابی روش پیشنهادی و مقایسه آن با روش‌های موجود، آزمایشات متعددی انجام شده‌است که در ادامه ضمن شرح هر یک از آزمایشات، نتایج آنها آمده‌است.

- **بررسی تاثیر مدل پیشنهادی بر دقت توصیه‌ها**

در این آزمایش برای بررسی میزان کارائی روش پیشنهادی و مقایسه آن با روش‌های موجود از معیار صحت توصیه‌ها و ترتیب توصیه‌ها استفاده شده‌است. در این آزمایش، روش پیشنهادی و روش‌های موجود با استفاده از مجموعه پایگاه داده‌ها (بخش ۴-۲) پیاده‌سازی و اجرا شده‌اند و پارامترهای (معیارهای) دقت، صحت (precision, recall)، معیار F1 ( F1 – measure)، و میزان درستی ترتیب منابع پیشنهادی ($DCG$) اندازه‌گیری می‌شود.

در آزمایش اول، نتایج اجرای روش پیشنهادی و الگوریتم $LOGO$ اندازه‌گیری شده‌اند. نتایج این آزمایش در جدول ۴-۲ آمده‌است. برای الگوریتم $LOGO$ پیاده‌سازی در دسترس عموم وجود ندارد و در این رساله بصورت کامل پیاده‌سازی شده‌است. این آزمایش با استفاده از پایگاه داده $NewsTweet$ انجام می‌شود. در این جدول، روش پیشنهادی با اختصار $RS$ نشان داده می‌شود.

جدول ۴-۲: نتایج اجرای روش پیشنهادی و الگوریتم LOGO.

| *top@N* | *method* | *precision* | *recall* | *F1 ± Variance* | *nDCG* |
|---|---|---|---|---|---|
| *5* | *LOGO* | 0.35 | 0.10 | 0.14 | 0.37 |
| | *RS* | 0.87 | 0.14 | 0.24 ± 0.0003 | 0.87 |
| *10* | *LOGO* | 0.32 | 0.16 | 0.19 | 0.44 |
| | *RS* | 0.88 | 0.29 | 0.43 ± 0.0009 | 0.87 |
| *15* | *LOGO* | 0.31 | 0.22 | 0.23 | 0.47 |
| | *RS* | 0.80 | 0.40 | 0.53 ± 0.0015 | 0.82 |
| *20* | *LOGO* | 0.31 | 0.28 | 0.27 | 0.49 |
| | *RS* | 0.75 | 0.50 | 0.59 ± 0.0025 | 0.78 |

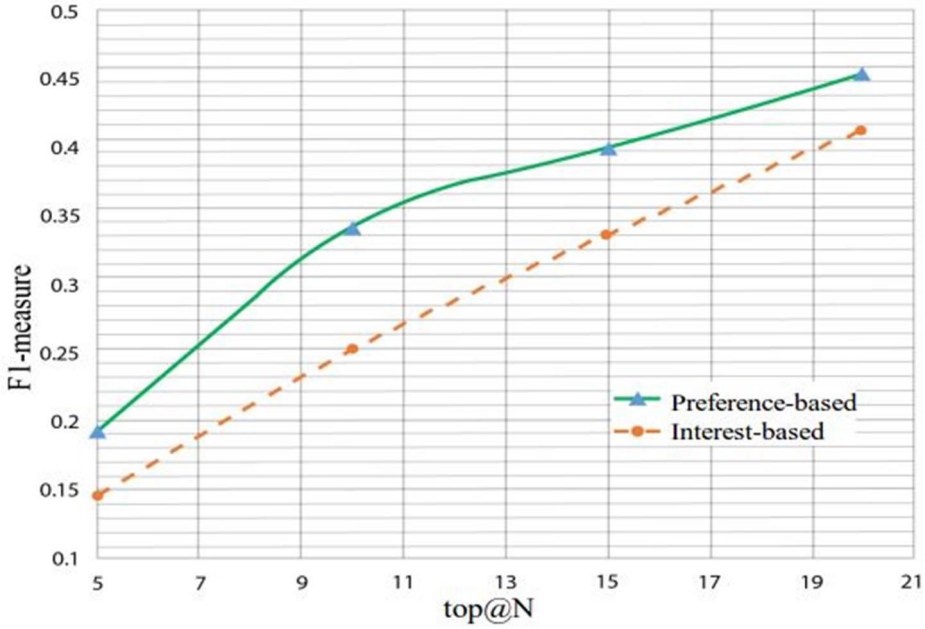
در آزمایش دوم نیز نتایج اجرای روش پیشنهادی و الگوریتم‌های پیاده‌سازی شده در ابزارهای *Librec*، *Surprise*، و *Lenskit* اندازه‌گیری شده‌است. در این آزمایش از پایگاه داده *MovieLens* استفاده شد. جدول ۴-۳ نتایج این آزمایش را نشان می‌دهد. مطابق این جدول، به عنوان نمونه در پنج پیشنهاد اول ($top@N = 5$)، روش پیشنهادی (*RS*) در معیار *DCG* نسبت به روش‌های موجود عملکرد بهتری دارد.

جدول ۴-۳: مقایسه نتایج اجرای روش پیشنهادی و ابزارهای معروف.

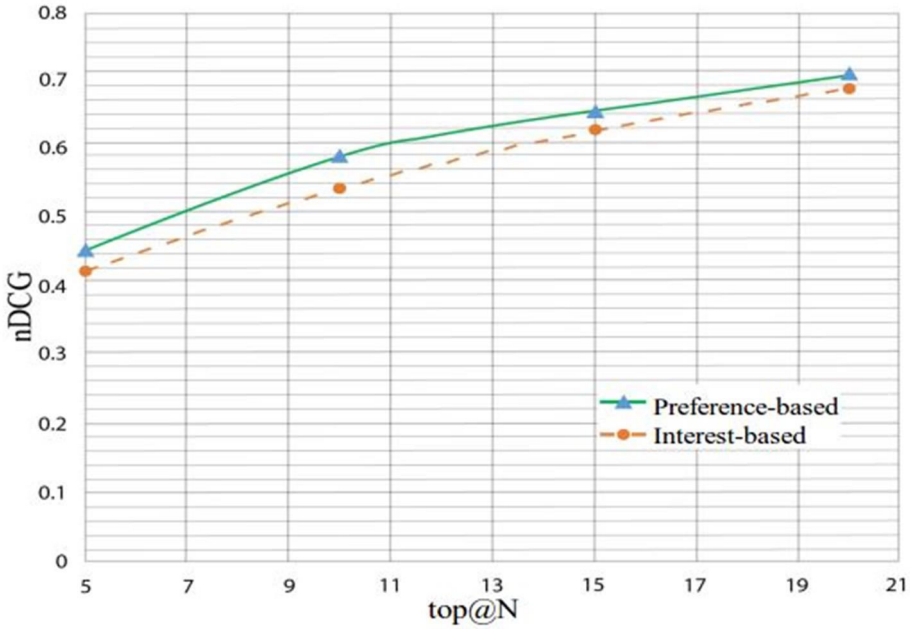
| *top@N* | *Method* | *Precision* | *nDCG* |
|---|---|---|---|
| *5* | *Lenskit: item-item* | 0.304 | 0.318 |
| | *Lenskit: funk-SVD* | 0.293 | 0.305 |
| | *Lenskit: contentbased* | 0.234 | 0.251 |
| | *Surprise: SVD* | 0.177 | 0.230 |
| | *Surprise: NMF* | 0.163 | 0.210 |
| | *Librec: timeSVD* | 0.234 | 0.251 |
| | *RS* | 0.753 | 0.461 |
| *10* | *Lenskit: item-item* | 0.462 | 0.492 |
| | *Lenskit: funk-SVD* | 0.454 | 0.481 |
| | *Lenskit: contentbased* | 0.399 | 0.423 |
| | *Surprise: SVD* | 0.347 | 0.409 |
| | *Surprise: NMF* | 0.328 | 0.383 |
| | *Librec: timeSVD* | 0.40 | 0.422 |
| | *RS* | 0.771 | 0.595 |
| *15* | *Lenskit: item-item* | 0.572 | 0.610 |
| | *Lenskit: funk-SVD* | 0.565 | 0.601 |
| | *Lenskit: contentbased* | 0.510 | 0.550 |
| | *Surprise: SVD* | 0.473 | 0.535 |
| | *Surprise: NMF* | 0.455 | 0.509 |
| | *Librec: timeSVD* | 0.510 | 0.550 |
| | *RS* | 0.689 | 0.656 |

### ◄ مقایسه مولفه‌های سلیقه و علاقه (تمایل) بر دقت توصیه‌ها

هدف از این آزمایش ارائه نتایج شهودی از مزیت بکارگیری ایده پیشنهادی این پژوهش یعنی بکارگیری مولفه **تمایل** در مدلسازی رفتار کار می‌باشد. برای ارائه نتایج شهودی، بکارگیری مدل پیشنهادی برای ایجاد سیستم توصیه‌گر با دو روش انجام می‌شود. در روش اول، ابتدا کاربر با استفاده از مولفه سلیقه مدل می‌شود، و سپس در یک سیستم توصیه‌گر بکارگرفته می‌شود. این روش اجر و نتایج اندازه‌گیری می‌شوند. در روش دوم، مدل رفتاری کاربر با دو مولفه **سلیقه و تمایل** (مدل پیشنهادی) ایجاد می‌شود، و سپس در یک سیستم توصیه‌گر بکارگرفته می‌شود. این روش نیز اجرا و نتایج اندازه‌گیری می‌شوند. نتایج این دو روش در شکل ۴-۱ آمده‌است. در این آزمایش از پایگاه داده *NewsTweet* استفاده شد. مطابق شکل ۴-۱، نتایج حاصل از این آزمایش بیان کننده کارائی روش پیشنهادی برای مدلسازی کاربر می‌باشد.




(a) F1-measure




شکل ۴-۱: بررسی تاثیر مدل پیشنهادی در کارائی سیستم توصیه‌گر و مقایسه دو روش مدلسازی "مبتنی بر سلیقه" و مبتنی بر زوج مولفه "سلیقه و تمایل".

### ◄ بررسی رفتار تکاملی مدل پیشنهادی

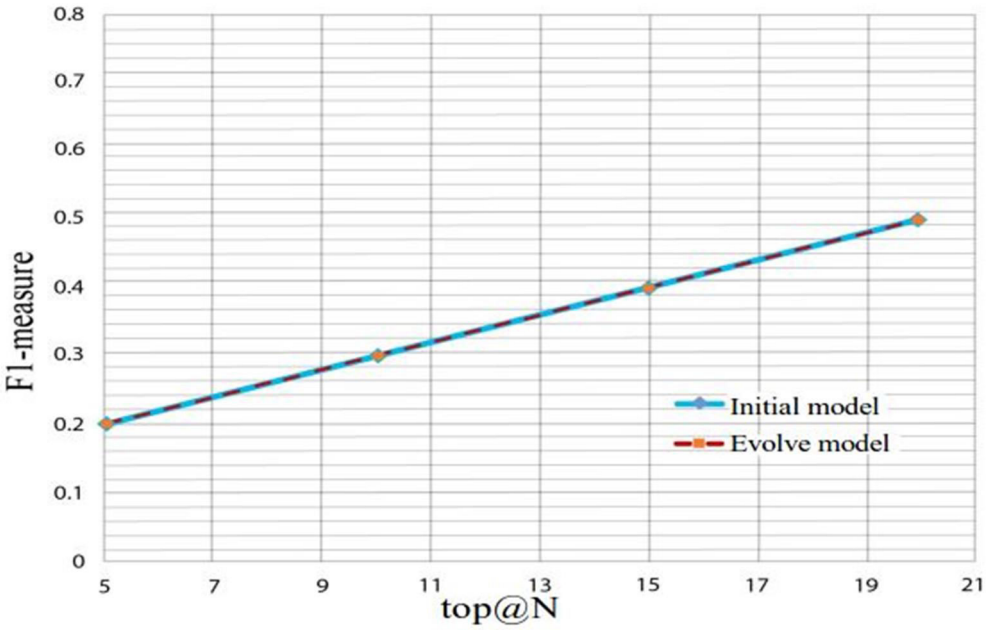
در این آزمایش میزان سازگاری روش پیشنهادی با انجام مشاهدات جدید بررسی می‌شود. برای بررسی رفتار مدل پیشنهادی در مواجهه با داده‌های جدید، ابتدا مجموعه داده (از پایگاه داده *MovieLens*) به دو قسمت تقسیم می‌شود. ابتدا قسمت اول برای ایجاد مدل و اندازه-گیری نتایج استفاده می‌شود. سپس قسمت دوم داده به مدل اعمال می‌شود و نتایج ارزیابی مجدد ارزیابی می‌شوند. مقایسه نتایج بدست آمده از این دو مرحله در شکل ۴-۲ آمده‌است. مطابق شکل ۴-۲، مدل پیشنهادی تا حد زیادی نسبت به مشاهدات جدید سازگار می‌باشد.

### ◄ تاثیر مدل پیشنهادی در بهبود تنکی داده

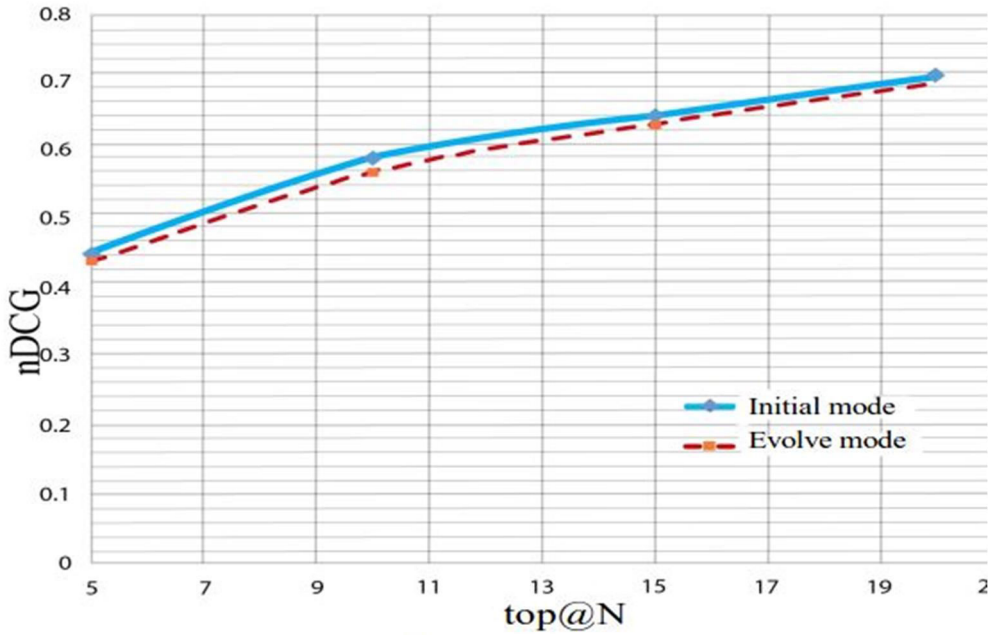
برای بررسی میزان کارائی روش پیشنهادی در مقابل مسئله تنکی داده، حدود ۳۵ درصد از پروفایل کاربران (به ازای هر کاربر) در پایگاه داده *MovieLens* بصورت تصادفی حذف شده و سپس روش پیشنهادی و روش‌های مورد ارزیابی بر روی این پایگاه داده تنک اعمال شدند. نتایج ارزیابی در جدول ۴-۴ آمده‌است. نتایج ارزیابی بیان کننده عملکرد بهتر روش پیشنهادی در مقابله با داده تنک می‌باشد.

جدول ۴-۴: مقایسه عملکرد روش پیشنهادی و روش‌های موجود در مواجه با مسئله تنکی داده.

| *top@N* | *Method* | *Precision* | *nDCG* |
|---|---|---|---|
| *3* | *Lenskit: item-item* | 0.149 | 0.154 |
| | *Lenskit: funk-SVD* | 0.142 | 0.141 |
| | *Surprise: NMF* | 0.046 | 0.071 |
| | *RS* | 0.615 | 0.398 |
| *6* | *Lenskit: item-item* | 0.234 | 0.274 |
| | *Lenskit: funk-SVD* | 0.230 | 0.239 |
| | *Surprise: NMF* | 0.120 | 0.151 |
| | *RS* | 0.607 | 0.503 |
| *9* | *Lenskit: item-item* | 0.298 | 0.320 |
| | *Lenskit: funk-SVD* | 0.292 | 0.311 |
| | *Surprise: NMF* | 0.177 | 0.211 |
| | *RS* | 0.616 | 0.586 |




(a) F1-measure

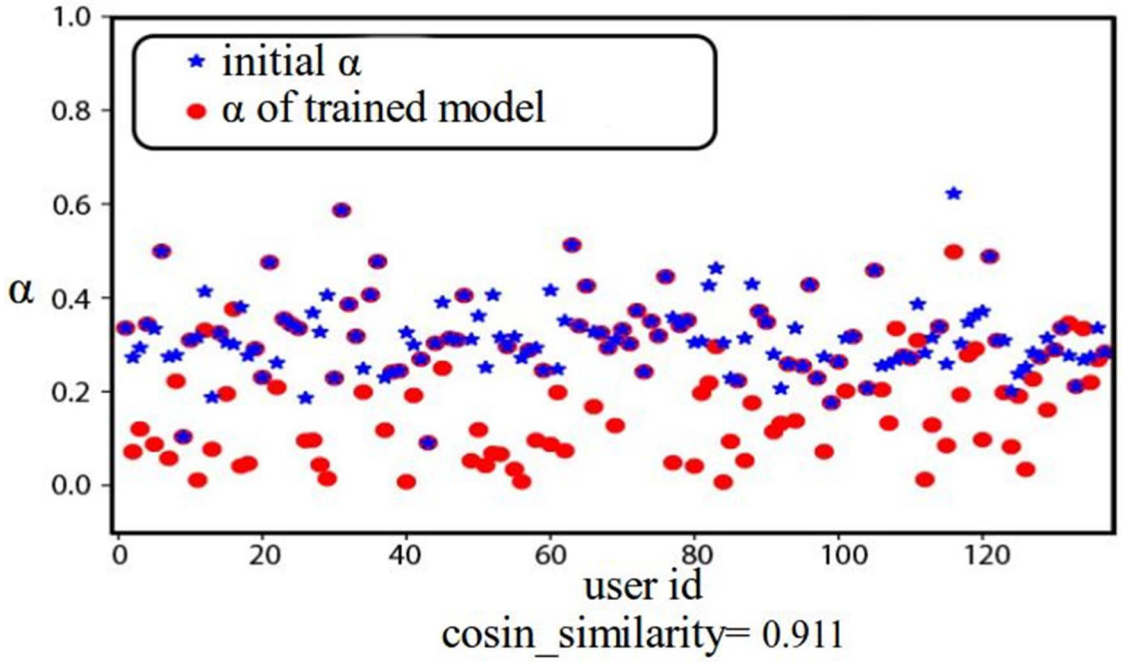



(b) DCG metric.

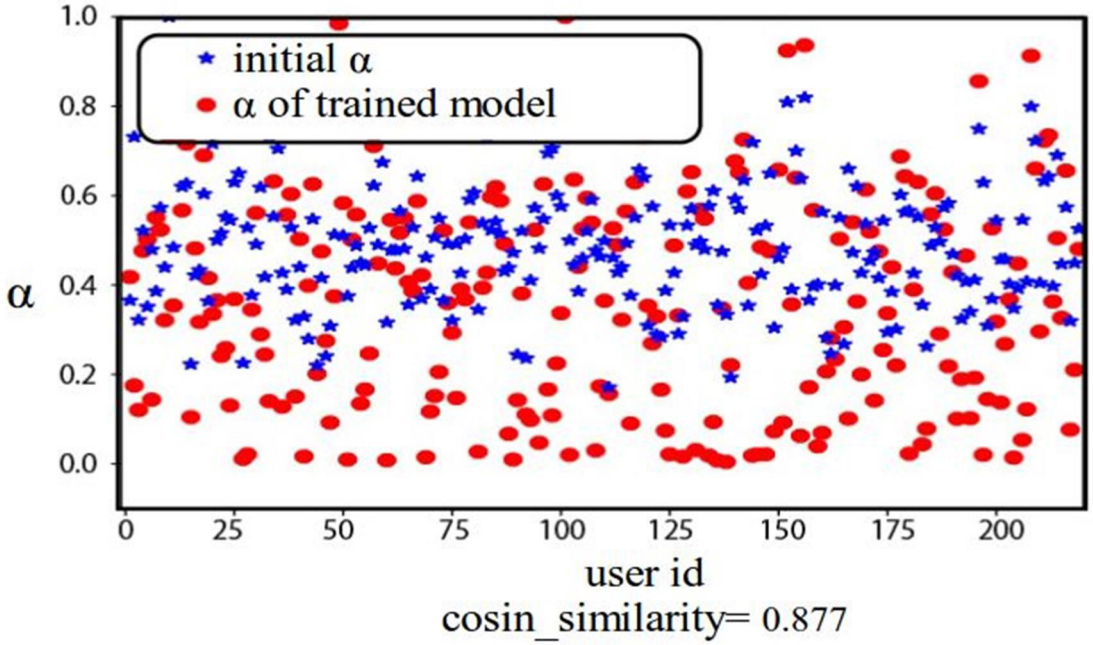
شکل۴-۲: ارزیابی سازگاری مدل پیشنهادی در مواجه با مشاهدات جدید.

### ۴-۴-۲- بررسی پارامتر توزیع سلایق کاربر

روش پیشنهادی نسبت به روش‌های موجود نظیر *timeSVD*، دارای کمترین تعداد پارامتر می‌باشد. در روش پیشنهادی، یک ابر-پارامتر برای مدل وجود دارد که پارامتر $\alpha$ می‌باشد. این پارامتر برای کنترل توزیع سلایق کاربر می‌باشد. مقدار کم آن سبب تولید تعداد سلایق زیاد می‌شود و بالعکس. در روش پیشنهادی برای تعیین مقدار اولیه این پارامتر از میانگین شباهت بین منابع انتخاب شده توسط هر کاربر، استفاده شده‌است. برای محاسبه میانگین شباهت منابع، ابتدا شباهت دو به دو همه منابع محاسبه شده است. سپس میانگین شباهت همه منابع به عنوان مقدار اولیه پارامتر $\alpha$ در نظر گرفته می‌شود. در این آزمایش (که با پایگاه داده *NewsTweet* انجام شد)، مقدار اولیه و مقادیر به روز شده پارامتر $\alpha$ در شکل ۴-۳ به تصویر کشیده شدند. این آزمایش، روش انتخاب شده در تعیین مقدار اولیه پارامتر $\alpha$ را تایید می‌کند.




(a)




(b)

شکل ۴-۳: نمایش همبستگی مقدار اولیه پارامتر توزیع سلایق با مقدار بهینه آن.

### ۴-۴-۳- بررسی تاثیر مولفه تمایل بر معیار شباهت در پالایش مشارکتی

در این آزمایش تاثیر مدل پیشنهادی در بهبود عملکرد سیستم‌های توصیه‌گر مبتنی بر پالایش مشارکتی بررسی می‌شود. این آزمایش نشان می‌دهد که استفاده از مدل پیشنهادی در معیار شباهت سیستم‌های توصیه‌گر مبتنی بر پالایش مشارکتی، تخمین مناسب‌تری در شناسایی کاربران مشابه و در نتیجه بهبود کارائی سیستم دارد. در این آزمایش (که با پایگاه داده *MovieLens* انجام شد)، روش پیشنهادی در یک الگوریتم ترکیبی برای ایجاد سیستم توصیه‌گر بکار گرفته شد. نتایج مقایسه روش پیشنهادی با روش‌های موجود (مبتنی بر پالایش مشارکتی) در جدول ۴-۵ آمده‌است. نتایج حاصل، بهبود عملکرد روش پیشنهادی را تایید می‌کند.

جدول ۴-۵: مقایسه عملکرد روش پیشنهادی با روش موجود مبتنی بر پالایش مشارکتی.

| *top@N* | *Method* | *Precision* | *nDCG* |
|---|---|---|---|
| *3* | *Lenskit: item-item* | 0.212 | 0.219 |
| | *Lenskit: funk-SVD* | 0.202 | 0.204 |
| | *Lenskit: contentbased* | 0.167 | 0.172 |
| | *Surprise: SVD* | 0.088 | 0.133 |
| | *Surprise: NMF* | 0.076 | 0.210 |
| | *Librec: timeSVD* | 0.142 | 0.158 |
| | *RS* | 0.746 | 0.521 |
| *6* | *Lenskit: item-item* | 0.342 | 0.360 |
| | *Lenskit: funk-SVD* | 0.331 | 0.346 |
| | *Lenskit: contentbased* | 0.267 | 0.285 |
| | *Surprise: SVD* | 0.217 | 0.274 |
| | *Surprise: NMF* | 0.203 | 0.253 |
| | *Librec: timeSVD* | 0.213 | 0.233 |
| | *RS* | 0.747 | 0.664 |
| *9* | *Lenskit: item-item* | 0.436 | 0.464 |
| | *Lenskit: funk-SVD* | 0.427 | 0.452 |
| | *Lenskit: contentbased* | 0.360 | 0.385 |
| | *Surprise: SVD* | 0.317 | 0.377 |
| | *Surprise: NMF* | 0.301 | 0.355 |
| | *Librec: timeSVD* | 0.280 | 0.300 |
| | *RS* | 0.751 | 0.775 |

## ۴-۵- جمع‌بندی

در این فصل آزمایشات متعددی برای ارزیابی میزان کارائی روش پیشنهادی انجام شده‌است. این آزمایشات سه جنبه مختلف روش‌های پیشنهادی را مورد ارزیابی قرار می‌دهند. این آزمایشات شامل ارزیابی صحت روش پیشنهادی، درستی ایده مدلسازی، و ارزیابی سیستم توصیه‌گر مبتنی بر روش پیشنهادی می‌باشد. در آزمایشات ارزیابی صحت، اندازه‌گیری پارامترهای دقت، بازخوانی( precision, recall, F1 – measure)، و میزان درستی ترتیب منابع پیشنهادی ($DCG$) و مقایسه آنها با روش‌های موجود نشان می‌دهد که روش پیشنهادی دارای عملکرد بهتری می‌باشد. در این پژوهش ایده اساسی در مدلسازی کاربر، استفاده از مولفه **تمایل** به همراه مولفه سلیقه می‌باشد. در آزمایش درستی ایده مدلسازی، معیارهای ارزیابی برای دو حالت "استفاده از مولفه سلیقه" و "استفاده از مولفه‌های سلیقه و تمایل"، اندازه‌گیری شدند. نتایج ارزیابی این آزمایش نشان می‌دهد که ایده استفاده شده در مدلسازی کاربر (مولفه تمایل) توانسته در بهبود عملکرد سیستم‌های توصیه‌گر موثر باشد. در آزمایش سوم نیز روش پیشنهادی در یک سیستم توصیه‌گر بکار گرفته شد. در ایجاد سیستم توصیه‌گر، از مدل پیشنهادی برای بهبود معیار شباهت استفاده می‌شود. در این آزمایش، اندازه‌گیری معیارهای ارزیابی نشان می‌دهد که روش پیشنهادی در بهبود معیار شباهت و بهبود عملکرد سیستم‌های توصیه‌گر ( در رویکرد پالایش مشارکتی) موثر می‌باشد.

# فصل پنجم

# جمع‌بندی و کارهای آینده

در این فصل جمع‌بندی پژوهش انجام شده در این رساله آمده‌است. همچنین پیشنهاداتی برای انجام تحقیقات و توسعه مدل ارائه شده در این رساله نیز بیان می‌شود.

## ۵-۱- جمع‌بندی پژوهش انجام شده در رساله

در این رساله رویکردها و روش‌های مختلف در توسعه سیستم‌های توصیه‌گر مورد بررسی قرار گرفته‌اند. با توجه به مطالعات انجام شده، رویکردهای مختلف در توسعه سیستم‌های توصیه‌گر را می‌توان به دو دسته رویکردهای مبتنی بر الگوریتم و رویکردهای مبتنی بر مدلسازی تقسیم نمود.

پالایش مشارکتی، پالایش محتوایی، و تکنیک‌های ترکیبی از جمله روش‌های اصلی رویکرد الگوریتمی در توسعه سیستم‌های توصیه‌گر می‌باشند. در پالایش مشارکتی، از پروفایل مجموعه‌ای از کاربران برای تولید منابع پیشنهادی استفاده می‌شود. به عبارتی در این روش با بکارگیری پروفایل همه کاربران، سلایق گروهی استخراج می‌شود. سپس به ازای درخواست یک کاربر، مجموعه افراد با سلایق مشترک شناسایی شده و براساس سلیقه گروهی، پیشنهاد انجام می‌شود. در پالایش محتوایی، از پروفایل هر کاربر برای ارائه توصیه استفاده می‌گردد. در این روش منابعی که به پروفایل کاربر شباهت بیشتری دارند، جهت توصیه به کاربر انتخاب می‌شوند. در این روش، غالبا اطلاعاتی از محتوای منابع استفاده شده وجود دارد. رویکرد ترکیبی نیز به ترکیب تکنیک‌ها و روش‌های مختلف برای ایجاد نتایج بهتر تمرکز دارد. به عنوان مثال در حوزه‌هایی که اطلاعات محتوایی منابع به همراه بازخورد کاربران موجود باشد می‌توان از رویکرد ترکیبی با مولفه‌های محتوایی و مشارکتی استفاده نمود.

در رویکرد مدلسازی پروفایل کاربر، هدف اصلی ایجاد مدل رفتاری مناسب از کاربر می‌باشد. در روش‌های موجود در زمینه بکارگیری پروفایل کاربران و مدلسازی آن، از مولفه سلایق کاربران استفاده شده‌است. در این روش‌ها با بکارگیری پروفایل کاربران و استخراج مولفه سلیقه (سلایق کاربران)، منابع مشابه با سلیقه (سلایق)، به کاربران پیشنهاد می‌شوند. براساس الگوریتم‌ها و تکنیک‌های اعمال شده بر پروفایل کاربران، سلیقه‌های استخراج شده ممکن است مبتنی بر فرد یا مبتنی بر گروهی از

کاربران باشند. به عنوان مثال در رویکرد پالایش مشارکتی عموما سلایق گروهی استخراج و استفاده می‌شود.

با مطالعات و بررسی‌های انجام شده در این رساله، **تمایل** کاربر به عنوان یک مولفه جدید در مدلسازی رفتار کاربران معرفی شده‌است. همچنانکه مولفه سلایق کاربر بیان کننده گروه‌هایی مشابه از منابع انتخاب شده توسط کاربر در طول زمان می‌باشد، مولفه تمایل کاربر، میزان اهمیت هر یک از سلایق کاربر در یک زمان معین را نشان می‌دهد. بدین ترتیب در این رساله با تحلیل پروفایل کاربران، رفتار کاربران با استفاده از یک بردار شامل زوج مولفه‌های <**سلیقه ، تمایل**> مدل می‌شود. سلایق کاربر نشان می‌دهد که کاربر در طول زمان چه گروه‌هایی از منابع را انتخاب کرده و تمایل کاربر، بیان کننده اهمیت هر سلیقه در زمان معین می‌باشد.

برای مدلسازی کاربر از چارچوب مدلسازی ترکیبی فرآیند دریکله (DPMM) استفاده شده‌است. این روش مدلسازی دارای دو مولفه می‌باشد. مولفه اول فرآیند خوشه‌بندی رستوران چینی (ddCRP) می‌باشد که امکان خوشه بندی افزایشی را فراهم می‌کند. بکارگیری این روش خوشه-بندی ضمن امکان استخراج سلیقه‌ها، امکان شناسایی سلیقه جدید را نیز فراهم می‌سازد. مولفه دوم، فرآیند دریکله می‌باشد که توزیع سلیقه‌ها (خوشه‌ها) را فراهم می‌سازد.

جهت بررسی صحت عملکرد و توانایی مدل پیشنهادی در مدلسازی رفتار کاربران در انتخاب منابع، ابتدا مدل پیشنهادی در یک سیستم توصیه‌گر محتوایی بکارگرفته شد. با پیاده‌سازی سیستم پیشنهادی و انجام آزمایشات، صحت مدل پیشنهادی تایید شده‌است. در ادامه، برای استفاده از مزایای رویکردهای پالایش مشارکتی و پالایش محتوایی، مدل پیشنهادی در یک سیستم توصیه‌گر ترکیبی بکارگرفته شد. در این سیستم ترکیبی، ابتدا با استفاده از مدل پیشنهادی در رویکرد مشارکتی، کاربران مشابه (براساس زوج مولفه <**سلیقه ، تمایل**>) بازیابی و سپس با رویکرد محتوایی، لیست مرتب شده‌ای از منابع کاندید به کاربر متقاضی ارائه می‌شود. نتایج آزمایش‌ها نشان می‌دهد که مدل

پیشنهادی توانسته عملکرد سیستم توصیه‌گر را بهبود دهد. همچنین آزمایشاتی جهت بررسی پویایی مدل پیشنهادی انجام شده که نتایج حاصل، سازگار مدل پیشنهادی همراه با تغییرات رفتار کاربر در طول زمان را تایید می‌کند.

## ۵-۲- کارهای آینده

با توجه به رویکردهای الگوریتمی و مدلسازی، در این رساله، رویکرد ایجاد مدل مناسب از رفتار کاربر با بکارگیری پروفایل کاربر تحقیق شده‌است. صرفنظر از رویکردها و روش‌های مختلفی که در توسعه سیستم‌های توصیه‌گر می‌توان ایجاد نمود، برای توسعه روش پیشنهادی، موارد ذیل را می‌توان تحقیق نمود.

- **استخراج سلایق کاربران.** در روش پیشنهادی برای استخراج سلایق کاربران از خوشه‌بند افزایشی ddCRP استفاده شده‌است. تحقیق در مورد بکارگیری خوشه‌بندهای دیگر نظیر Indian Buffet Process (IBP) و خوشه‌بند سلسله مراتبی به عنوان یکی از موضوعات مناسب در بهبود عملکرد مدل پیشنهادی می‌باشد. به عنوان مثال هنگامی‌که بتوان رفتار کاربران در انتخاب منابع را با استفاده از ماتریس بازخورد (که سطرها ستون‌ها، و درایه‌ها به ترتیب کاربران، منابع، و امتیاز کاربر به منبع نظیر را نشان می‌دهند) گردآوری نمود، استفاده از فرآیند خوشه‌بندی بوفه هندی (IBP) مدل مناسبی از رفتار کاربران در ثبت بازخوردها ایجاد می‌کند. در مدل پیشنهادی، در تحلیل پروفایل هر کاربر و ایجاد خوشه‌بندی، هر منبع به یک خوشه (سلیقه) تعلق دارد. بنابراین استفاده از مفاهیم فازی در خوشه‌بندی نیز قابل توجه می‌باشد.
- **استنتاج تمایلات کاربران.** در روش پیشنهادی برای محاسبه تمایلات کاربران از دو مولفه زمان و تعداد تعاملات در هر خوشه استفاده شده‌است. برای بهبود عملکرد روش پیشنهادی، توسعه‌هایی را می‌توان در محاسبه تمایلات کاربران تحقیق نمود. به عنوان مثال در حوزه‌های کاری که امتیاز کاربران به منابع موجود باشد، می‌توان مولفه امتیاز را نیز در محاسبه تمایلات کاربر در نظر گرفت.

همچنین استفاده از توابع توزیع استاندارد نظیر توزیع‌های وابسته به زمان (پواسن) را می‌توان مورد بررسی قرار داد.

- **الگوریتم ترکیبی.** در این رساله مدل پیشنهادی کاربر در یک مدل ترکیبی بکار گرفته شد. روش‌های مختلفی برای توسعه الگوریتم ترکیبی وجود دارند که می‌توان از آنها برای بهبود سیستم توصیه‌گر پیشنهادی (در این رساله) استفاده نمود. علاوه براین، استفاده از چارچوب شبکه بیزین و ایجاد یک مدل گرافیکی احتمالاتی چند سطحی (Hierarchical Probabilistic Graphical Model) به عنوان یک زمینه مناسب تحقیقاتی برای بهبود روش پیشنهادی می‌باشد. در مدل احتمالاتی چند سطحی، سطح اول، پروفایل کل کاربران را در نظر می‌گیرد و سطوح داخلی با در نظر گرفتن پروفایل هر کاربر، مدلسازی رفتاری کاربران را انجام می‌دهد. بدین ترتیب، الگوریتم ترکیبی در قالب مدل گرافیکی احتمالاتی چند سطحی پیاده‌سازی می‌شود.

- **حفظ حریم خصوصی (privacy) در توسعه سیستم‌های توصیه‌گر**

سیستم‌های توصیه‌گر جهت ایجاد یک محیط شخصی‌سازی شده، از پروفایل کاربران استفاده می‌کنند. این پروفایل شامل داده‌هایی نظیر اطلاعات شخصی کاربران، علایق آنها، موقعیت مکانی می‌باشد. از اینرو حفظ حریم خصوصی کاربران یکی از مسائل اساسی در توسعه و بکارگیری سیستم‌های توصیه‌گر می‌باشد. فروش و یا فاش شدن اطلاعات از جمله تهدیداتی است که اعتماد استفاده از سیستم‌های توصیه‌گر را تحت تاثیر قرار می‌دهد. این تهدیدات ممکن است از طرف خود سیستم توصیه‌گر، دیگر کاربران سیستم، و نهادهای خارجی (هکرها) ایجاد شود. در جهت حفظ حریم خصوصی، عموما سه رویکرد وجود دارد:

**معماری، زیرساخت، و استانداردها.** این رویکرد تاکید بر حداقل سازی اشتراک و متمرکز سازی پروفایل کاربران دارد. معماری توزیع شده، امکان جابجایی پروفایل کاربران، و الگوریتم سمت کاربر از جمله راهکارهای این رویکرد می‌باشند.

**الگوریتم**. در این رویکرد از تکنیک‌های محافظت از داده استفاده می‌شود. به عنوان مثال، ناشناس نمودن شناسه کاربران، رمزنگاری، و مبهم‌سازی از جمله راهکارهای این رویکرد می‌باشند.

**سیاست‌گذاری**. این رویکرد غالبا غیر تکنیکی می‌باشد. این رویکرد شامل ایجاد قواعد و مقرارتی می‌باشد که کاربران و فراهم‌کنندگان خدمات (با نظارت سازمان‌های معتبر) باید آنرا رعایت نمایند. به عنوان مثال در وب سایت‌هایی مانند *Yahoo* و *Google*، هر کاربر برای استفاده از خدمات باید ابتدا انتخاب " I agree to the Yahoo Terms and Privacy" را تایید نمایند.

در جهت تامین حریم خصوصی کاربران، عموما از روش‌های تکنیکی برای توسعه سیستم‌های توصیه‌گر استفاده می‌شود.

- **استنتاج فرآیند تولید منابع پیشنهادی توسط سیستم‌های توصیه‌گر**

یکی از روش‌های ارزیابی منابع توصیه شده توسط سیستم توصیه‌گر، وجود فرآیند تولید منابع می‌باشد. به عبارتی دلیل انتخاب هر یک از منابع پیشنهاد شده معین باشد. به عنوان مثال در رویکرد پالایش مشارکتی، میزان تاثیر و تعداد افراد مشابهی که منبع مورد نظر را پیشنهاد داده‌اند گزارش شود. در رویکرد مبتنی بر پایگاه دانش، مراحل استنتاج و اعمال قوانین و بکارگیری دانش در جهت تولید منابع پیشنهادی، به کاربر متقاضی ارائه شود.

- **بکارگیری سیستم‌های توصیه‌گر در حوزه‌های دیگر**

بکارگیری سیستم‌های توصیه‌گر در حوزه‌های مختلف نظیر صنعت تولید نرم‌افزار، آموزش الکترونیک، تولید دارو در حال توسعه می‌باشد. بدین ترتیب، نحوه بکارگیری روش‌ها و مدل‌های موجود، جهت بومی‌سازی و توسعه یک سیستم توصیه‌گر در این حوزه‌ها، از مسائل قابل توجه

می‌باشد. به عنوان مثال از رویکرد پالایش مشارکتی می‌توان در حوزه تولید دارو استفاده نمود [۱۰۲, ۱۰۳]. در این حوزه، جهت تولید دارو نیازمند انجام آزمایشات متعدد جهت بررسی یک ترکیب دارویی بر پروتئین‌های مختلف می‌باشد. با استفاده از نتایج آزمایشات گذشته، می‌توان ماتریس بازخورد را ایجاد نمود که سطرها و ستون‌ها به ترتیب ترکیبات دارویی و پروتئین‌ها را نشان می‌دهند. این ماتریس بازخورد در پالایش مشارکتی قابل استفاده می‌باشد. هر درایه از این ماتریس بیان کننده نتیجه اعمال یک ترکیب داروی بر یک پروتئین می‌باشد.

بدین ترتیب نحوه مدلسازی هر حوزه کاری جهت بکارگیری سیستم‌های توصیه‌گر به یکی دیگر از مسائل قابل توجه در توسعه سیستم‌های توصیه‌گر می‌باشد.

## فهرست واژگان

**B**

Bayesian non-Parametric framework چارچوب ناپارامتریک بیزی

**C**

Cosine تابع محاسبه شباهت

Chinese Restaurant Process فرآیند رستوران چینی

Changing User Interest تغییرات سلیقه کاربر

Content Based Filtering پالایش مشارکتی

Cold Start شروع سرد

Collaborative filtering پالایش مشارکتی

Content Based Filtering پالایش مبتنی بر محتوا

**D**

Dirichlet Process Mixture Model مدل ترکیبی فرآیند دریکله

Dirichlet Process فرآیند دریکله

Demographics اطلاعات فردی

Distance Dependent Chienece Restaurant Process فرآیند رستوران چینی مبتنی بر فاصله

Discounted cumulative gain معیار ارزیابی کیفیت رتبه بندی

**E**

ensemble hierarchical clustering ادغام خوشه‌بندی سلسله مراتبی

Explicit صریح

Expectation Maximization بیشینه‌سازی امید

**F**

Feedback Matrix ماتریس بازخورد

Feature Augmentation هم افزایی ویژگی

F1-measure معیار ترکیبی ارزیابی

**H**

Hierarchical Probabilistic Graphical مدل گرافیکی احتمالاتی چند سطحی

**I**

Implicit ضمنی

Indian Buffet Process فرآیند بوفه هندی

**J**

Jaccard تابع محاسبه شباهت

**L**

Latent factor model مدل فاکتور پنهان

**M**

Memory-based مبتنی بر حافظه

Model-based مبتنی بر مدل

mean square difference میانگین مربعات خطا

Matrix-Factorization تجزیه ماتریس

Memory-based Collaborative Filtering پالایش مشارکتی مبتنی بر حافظه

**N**

Nearest Neighbor نزدیکترین همسایگی

**P**

Probabilistic Graphical Model مدل گرافیکی احتمالاتی

Preference تمایل

Precision دقت

Pearson تابع محاسبه شباهت

Privacy حریم خصوصی

posterior inference استنتاج ثانویه

**R**

Recall صحت

**S**

Sparsity تنکی

Stochastic Gradient Descent کاهش شیب تصادفی

**T**

Trust اعتماد

Topic Distribution توزیع موضوعات

**U**

Unpredictable Items منابع با محتوای غیرقابل پیش‌بینی

# مراجع


1. Linden, G., B. Smith, and J. York, (2003) "*Amazon. com recommendations: Item-to-item collaborative filtering".* IEEE Internet computing, (1): p. 76-80.
2. Adomavicius, G. and Y. Kwon, (2012) "*Improving aggregate recommendation diversity using ranking-based techniques".* IEEE Transactions on Knowledge and Data Engineering, **24**(5): p. 896-911.
3. Aggarwal, C.C., (2016) "*An introduction to recommender systems*, in *Recommender Systems*", Springer. p. 1-28.
4. Kotkov, D., S. Wang, and J. Veijalainen, (2016) "*A survey of serendipity in recommender systems.* Knowledge-Based Systems", **111**: p. 180-192.
5. Ricci, F., L. Rokach, and B. Shapira, (2015) "*Recommender systems: introduction and challenges*, in *Recommender systems handbook*", Springer. p. 1-34.
6. Bobadilla, J., Bobadilla, J., Ortega, F., Hernando, A. and Gutiérrez, A., (2013) "*Recommender systems survey.* Knowledge-based systems", **46**: p. 109-132.
7. Adomavicius, G. and A. Tuzhilin, (2005) "*Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions".* IEEE Transactions on Knowledge & Data Engineering, (6): p. 734-749.
8. Koren, Y., R. Bell, and C. Volinsky, (2009) "*Matrix factorization techniques for recommender systems".* Computer, (8): p. 30-37.
9. Schafer, J.B., Frankowski, D., Herlocker, J. and Sen, S., (2007) "*Collaborative filtering recommender systems"*, in *The adaptive web*, Springer. p. 291-324.
10. Pazzani, M.J. and D. Billsus, (2007) "*Content-based recommendation systems"*, in *The adaptive web*, Springer. p. 325-341.
11. Martín-Blas, T. and A. Serrano-Fernández, (2009) "*The role of new technologies in the learning process: Moodle as a teaching tool in Physics".* Computers & Education, **52**(1): p. 35-44.
12. Frost, J.H. and M.P. Massagli, (2008) "*Social uses of personal health information within PatientsLikeMe, an online patient community: what can happen when patients have access to one another's data".* Journal of medical Internet research, **10**(3).
13. Schafer, J.B., J. Konstan, and J. Riedl. (1999) "*Recommender systems in e-commerce".* in *Proceedings of the 1st ACM conference on Electronic commerce*, ACM. pp. 158-166.
14. Schafer, J.B., J.A. Konstan, and J. Riedl, (2001) "*E-commerce recommendation applications".* Data mining and knowledge discovery, **5**(1-2): p. 115-153.
15. Lu, J., Wu, D., Mao, M., Wang, W. and Zhang, G., (2015) "*Recommender system application developments: a survey".* Decision Support Systems, **74**: p. 12-32.
16. Jung, S., J.L. Herlocker, and J. Webster, (2007) "*Click data as implicit relevance feedback in web search".* Information Processing & Management, **43**(3): p. 791-807.
17. Kumar, M., Yadav, D.K., Singh, A. and Gupta, V.K., (2015) "*A movie recommender system: Movrec".* International Journal of Computer Applications, **124**(3).

18. Gomez-Uribe, C.A. and N. Hunt, (2016) "*The netflix recommender system: Algorithms, business value, and innovation".* ACM Transactions on Management Information Systems (TMIS), **6**(4): p. 13.
19. Aggarwal, C.C., (2016) "*Recommender systems".* Springer.
20. Wu, H., Pei, Y., Li, B., Kang, Z., Liu, X. and Li, H., (2015) "*Item recommendation in collaborative tagging systems via heuristic data fusion".* Knowledge-Based Systems, **75**: p. 124-140.
21. Zheng, N. and Q. Li, (2011) "*A recommender system based on tag and time information for social tagging systems".* Expert Systems with Applications, **38**(4): p. 4575-4587.
22. Lops, P., M. De Gemmis, and G. Semeraro, (2011) "*Content-based recommender systems: State of the art and trends"*, in *Recommender systems handbook*, Springer. p. 73-105.
23. Palmisano, C., A. Tuzhilin, and M. Gorgoglione, (2008) "*Using context to improve predictive modeling of customers in personalization applications".* IEEE transactions on knowledge and data engineering, **20**(11): p. 1535-1549.
24. Carrer-Neto, W., Hernández-Alcaraz, M.L., Valencia-García, R. and García-Sánchez, F., (2012) "*Social knowledge-based recommender system. Application to the movies domain".* Expert Systems with applications, **39**(12): p. 10990-11000.
25. Burke, R., (2002) "*Hybrid recommender systems: Survey and experiments".* User modeling and user-adapted interaction, **12**(4): p. 331-370.
26. Gauch, S., Speretta, M., Chandramouli, A. and Micarelli, A., (2007) "*User profiles for personalized information access"*, in *The adaptive web*. Springer. p. 54-89.
27. Jawaheer, G., P. Weller, and P. Kostkova, (2014) "*Modeling user preferences in recommender systems: A classification framework for explicit and implicit user feedback".* ACM Transactions on Interactive Intelligent Systems (TiiS), **4**(2): p. 8.
28. Symeonidis, P., A. Nanopoulos, and Y. Manolopoulos. (2007) "*Feature-weighted user model for recommender systems".* in *International Conference on User Modeling*. Springer.
29. Xu, K., Zheng, X., Cai, Y., Min, H., Gao, Z., Zhu, B., Xie, H. and Wong, T.L., (2018) "*Improving user recommendation by extracting social topics and interest topics of users in uni-directional social networks".* Knowledge-Based Systems, **140**: p. 120-133.
30. Yin, H., Cui, B., Chen, L., Hu, Z. and Zhou, X., (2015) "*Dynamic user modeling in social media systems".* ACM Transactions on Information Systems (TOIS), **33**(3): p. 10.
31. Brusilovsky, P. and E. Millán, (2007) "*User models for adaptive hypermedia and adaptive educational systems"*, in *The adaptive web*, Springer. p. 3-53.
32. Cai, Y., Leung, H.F., Li, Q., Min, H., Tang, J. and Li, J., (2014) "*Typicality-based collaborative filtering recommendation".* IEEE Transactions on Knowledge and Data Engineering, **26**(3): p. 766-779.
33. Koren, Y. (2009) "*Collaborative filtering with temporal dynamics".* in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM. pp. 447-456.
34. Aggarwal, C.C., (2016) "*Content-based recommender systems"*, in *Recommender systems*, Springer. p. 139-166.

35. Sahoo, N., P.V. Singh, and T. Mukhopadhyay, (2012) "*A hidden Markov model for collaborative filtering".* Mis Quarterly, p. 1329-1356.
36. Das, A.S., Datar, M., Garg, A. and Rajaram, S. (2007) "*Google news personalization: scalable online collaborative filtering".* in *Proceedings of the 16th international conference on World Wide Web*. ACM. pp. 271-280.
37. Luo, X., Zhou, M., Xia, Y. and Zhu, Q., (2014) "*An efficient non-negative matrix-factorization-based approach to collaborative filtering for recommender systems".* IEEE Transactions on Industrial Informatics, **10**(2): p. 1273-1284.
38. Spangler, K.L., (1983) "*Reading interests vs. reading preferences: using the research".* The Reading Teacher, **36**(9): p. 876-878.
39. Blei, D.M. and P.I. Frazier, (2011) "*Distance dependent Chinese restaurant processes".* Journal of Machine Learning Research, **12**(Aug): p. 2461-2488.
40. Teh, Y.W., Jordan, M.,Beal, M., Blei, M., (2006) "*Hierarchical dirichlet processes".* Journal of the American Statistical Association, **101**(476), pp. 1566-1581
41. Gershman, S.J. and D.M. Blei, (2012) "*A tutorial on Bayesian nonparametric models".* Journal of Mathematical Psychology, **56**(1): p. 1-12.
42. Avazpour, I., Pitakrat, T., Grunske, L. and Grundy, J., (2014) "*Dimensions and Metrics for Evaluating Recommendation Systems"*, in *Recommendation Systems in Software Engineering*, M.P. Robillard, et al., Editors., Springer Berlin Heidelberg: Berlin, Heidelberg. p. 245-273.
43. Bagher, R.C., H. Hassanpour, and H. Mashayekhi, (2017) "*User trends modeling for a content-based recommender system".* Expert Systems With Applications, **87**: p. 209-219.
44. Rana, C. and S.K. Jain, (2014) "*An evolutionary clustering algorithm based on temporal features for dynamic recommender systems".* Swarm and Evolutionary Computation, **14**: p. 21-30.
45. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J., (1994) "*GroupLens: an open architecture for collaborative filtering of netnews".* in *Proceedings of the 1994 ACM conference on Computer supported cooperative work*. ACM. pp. 175-186.
46. Deshpande, M. and G. Karypis, (2004) "*Item-based top-n recommendation algorithms".* ACM Transactions on Information Systems (TOIS), **22**(1): p. 143-177.
47. Koren, Y. and R. Bell, (2015) "*Advances in collaborative filtering*, in *Recommender systems handbook*", Springer. p. 77-118.
48. Zhao, Z.-D. and M.-S. Shang. (2010) "*User-based collaborative-filtering recommendation algorithms on hadoop".* in *Knowledge Discovery and Data Mining, 2010. WKDD'10. Third International Conference on*.. IEEE. pp. 478-481.
49. Melville, P., R.J. Mooney, and R. Nagarajan, (2002) "*Content-boosted collaborative filtering for improved recommendations".* Aaai/iaai, **23**: p. 187-192.
50. Patra, B.K., Launonen, R., Ollikainen, V. and Nandi, S., (2015) "*A new similarity measure using Bhattacharyya coefficient for collaborative filtering in sparse data".* Knowledge-Based Systems, **82**: p. 163-177.
51. Liu, H., Hu, Z., Mian, A., Tian, H. and Zhu, X., (2014) "*A new user similarity model to improve the accuracy of collaborative filtering".* Knowledge-Based Systems, **56**: p. 156-166.

52. Goldberg, D., Nichols, D., Oki, B.M. and Terry, D., (1992) "*Using collaborative filtering to weave an information tapestry".* Communications of the ACM, **35**(12): p. 61-70.
53. Breese, J.S., D. Heckerman, and C. Kadie. (1998) "*Empirical analysis of predictive algorithms for collaborative filtering".* in *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*.. Morgan Kaufmann Publishers Inc. pp. 43-52.
54. Konstan, J.A., Miller, B.N., Maltz, D., Herlocker, J.L., Gordon, L.R. and Riedl, J., *(*1997) "*GroupLens: applying collaborative filtering to Usenet news".* Communications of the ACM, **40**(3): p. 77-87.
55. Sarwar, B., Karypis, G., Konstan, J. and Riedl, J., (2001) "*Item-based collaborative filtering recommendation algorithms".* in *Proceedings of the 10th international conference on World Wide Web*. ACM. pp. 285-295.
56. Mobasher, B., R. Burke, and J.J. Sandvig., (2006) "*Model-based collaborative filtering as a defense against profile injection attacks".* in *AAAI*. p. 1388.
57. Robles, V., Larranaga, P., Menasalvas, E., Pérez, M.S. and Herves, V., (2003 ) "*Improvement of Naive Bayes collaborative filtering using interval estimation".* in *Web Intelligence, 2003. WI 2003. Proceedings. IEEE/WIC International Conference on*. IEEE. pp. 168-174.
58. Yang, X., Y. Guo, and Y. Liu, (2013) "*Bayesian-inference-based recommendation in online social networks".* IEEE Transactions on Parallel and Distributed Systems, **24**(4): p. 642-651.
59. Miyahara, K. and M.J. Pazzani, (2000) "*Collaborative filtering with the simple Bayesian classifier".* in *Pacific Rim International conference on artificial intelligence*. Springer. pp. 679-689.
60. Wang, K. and Y. Tan. (2011) "*A new collaborative filtering recommendation approach based on naive Bayesian method".* in *International Conference in Swarm Intelligence*. Springer. pp. 218-227.
61. Jung, K.-Y., H.-J. Hwang, and U.-G. Kang. (2006) "*Constructing full matrix through naive Bayesian for collaborative filtering".* in *International Conference on Intelligent Computing*. Springer. pp. 1210-1215.
62. Ghazanfar, M. and A. Prugel-Bennett, (2010) "*An improved switching hybrid recommender system using Naive Bayes classifier and collaborative filtering".* In Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval .pp. 39-46.
63. Ono, C., Kurokawa, M., Motomura, Y. and Asoh, H., (2007) "*A context-aware movie preference model using a Bayesian network for recommendation and promotion".* in *International Conference on User Modeling*. Springer. pp. 247-257.
64. Park, M.-H., J.-H. Hong, and S.-B. Cho. (2007) "*Location-based recommendation system using bayesian user's preference model in mobile devices".* in *International Conference on Ubiquitous Intelligence and Computing*. Springer. pp. 1130-1139.
65. Ashkezari-T, S. and M.-R. Akbarzadeh-T. (2010) "*Fuzzy-Bayesian network approach to genre-based recommender systems".* in *Fuzzy Systems (FUZZ), 2010 IEEE International Conference on*. IEEE. pp. 1-7.
66. Yu, K., Schwaighofer, A., Tresp, V., Xu, X. and Kriegel, H.P., (2004) "*Probabilistic memory-based collaborative filtering".* IEEE Transactions on Knowledge and Data Engineering, **16**(1): p. 56-69.

67. Hofmann, T. (2003) "*Collaborative filtering via gaussian probabilistic latent semantic analysis*". in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*. ACM. pp. 259-266.
68. Zhang, Y. and J. Koren. (2007) "*Efficient bayesian hierarchical user modeling for recommendation system*". in *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM. pp. 47-54.
69. Ungar, L.H. and D.P. Foster. (1998) "*Clustering methods for collaborative filtering*". in *AAAI workshop on recommendation systems*. pp. 114-129.
70. Ma, H., Zhou, T.C., Lyu, M.R. and King, I., (2011) "*Improving recommender systems by incorporating social contextual information*". ACM Transactions on Information Systems (TOIS), **29**(2): p. 9.
71. Terveen, L. and W. Hill, (2001) "*Beyond recommender systems: Helping people help each other*". HCI in the New Millennium, **1**(2001): p. 487-509.
72. Felfernig, A., Zhou, T.C., Lyu, M.R. and King, I., (2014) "*Basic approaches in recommendation systems*", in *Recommendation Systems in Software Engineering*, Springer. p. 15-37.
73. Li, L., Wang, D., Li, T., Knox, D. and Padmanabhan, B., (2011) "*SCENE: a scalable two-stage personalized news recommendation system*". in *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. ACM. pp. 125-134.
74. Li, L., Zheng, L., Yang, F. and Li, T., (2014) "*Modeling and broadening temporal user interest in personalized news recommendation*". Expert Systems with Applications, **41**(7): p. 3168-3177.
75. Krulwich, B., (1997) "*Lifestyle finder: Intelligent user profiling using large-scale demographic data*". AI magazine, **18**(2): p. 37.
76. Wang, Y., S.C.-F. Chan, and G. Ngai. (2012) "*Applicability of demographic recommender system to tourist attractions: a case study on trip advisor*". in *Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 03*. IEEE Computer Society. pp. 97-101.
77. Pazzani, M.J., (1999) "*A framework for collaborative, content-based and demographic filtering*". Artificial intelligence review, **13**(5-6): p. 393-408.
78. Kautz, H., B. Selman, and M. Shah, *Referral Web: combining social networks and collaborative filtering.* Communications of the ACM, 1997. **40**(3): p. 63-65.
79. Kautz, H., B. Selman, and A. Milewski. (1996) "*Agent amplified communication*". in *AAAI/IAAI, Vol. 1*. pp. 3-9.
80. Tang, J., X. Hu, and H. Liu, (2013) "*Social recommendation: a review*". Social Network Analysis and Mining, **3**(4): p. 1113-1133.
81. O'Donovan, J. and B. Smyth. (2005) "*Trust in recommender systems*". in *Proceedings of the 10th international conference on Intelligent user interfaces*. ACM. pp. 17-24.
82. Massa, P. and P. Avesani. (2004) "*Trust-aware collaborative filtering for recommender systems*". in *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems*". Springer. pp. 492-508.
83. Horowitz, D. and S.D. Kamvar. (2010) "*The anatomy of a large-scale social search engine*". in *Proceedings of the 19th international conference on World wide web*. ACM. pp. 431-440.

84. Hariri, N., B. Mobasher, and R. Burke. (2014) "*Context adaptation in interactive recommender systems".* in *Proceedings of the 8th ACM Conference on Recommender systems*. ACM. pp. 41-48.
85. Otebolaku, A.M. and M.T. Andrade, (2014) "*A context-aware framework for media recommendation on smartphones"*, in *ECUMICT 2014*, Springer. p. 87-108.
86. Adomavicius, G., Sankaranarayanan, R., Sen, S. and Tuzhilin, A., (2005) "*Incorporating contextual information in recommender systems using a multidimensional approach".* ACM Transactions on Information Systems (TOIS), **23**(1): p. 103-145.
87. Barranco, M.J., Noguera, J.M., Castro, J. and Martínez, L., (2012) "*A context-aware mobile recommender system based on location and trajectory*, in *Management intelligent systems"*, Springer. p. 153-162.
88. Trewin, S., (2000) "*Knowledge-based recommender systems".* Encyclopedia of library and information science, **69**(Supplement 32): p. 180.
89. Felfernig, A. and R. Burke, (2008) "*Constraint-based recommender systems: technologies and research issues".* in *Proceedings of the 10th international conference on Electronic commerce*. ACM. p. 3.
90. Balabanović, M. and Y. Shoham, (1997) "*Fab: content-based, collaborative recommendation".* Communications of the ACM, **40**(3): p. 66-72.
91. Burke, R., (2007) "*Hybrid web recommender systems"*, in *The adaptive web*, Springer. p. 377-408.
92. Zheng, L., Li, L., Hong, W. and Li, T., (2013) "*PENETRATE: Personalized news recommendation using ensemble hierarchical clustering".* Expert Systems with Applications, **40**(6): p. 2127-2136.
93. Blei, D.M., A.Y. Ng, and M.I. Jordan, (2003) "*Latent dirichlet allocation".* Journal of machine Learning research, **3**(Jan): p. 993-1022.
94. Schafer, J.B., (2009) "*The Application of Data-Mining to Recommender Systems"*, in *Encyclopedia of Data Warehousing and Mining, Second Edition*, W. John, Editor., IGI Global: Hershey, PA, USA. p. 45-50.
95. Kim, H.-N., Alkhaldi, A., El Saddik, A. and Jo, G.S., (2011) "*Collaborative user modeling with user-generated tags for social recommender systems".* Expert Systems with Applications, **38**(7): p. 8488-8496.
96. Chen, C. and J. Ren, (2017) "*Forum latent Dirichlet allocation for user interest discovery".* Knowledge-Based Systems, **126**: p. 1-7.
97. Herlocker, J.L., Konstan, J.A., Terveen, L.G. and Riedl, J.T., (2004) "*Evaluating collaborative filtering recommender systems".* ACM Transactions on Information Systems (TOIS), **22**(1): p. 5-53.
98. Del Olmo, F.H. and E. Gaudioso, (2008) "*Evaluation of recommender systems: A new approach".* Expert Systems with Applications, **35**(3): p. 790-804.
99. Görür, D. and C.E. Rasmussen, (2010) "*Dirichlet process gaussian mixture models: Choice of the base distribution".* Journal of Computer Science and Technology, **25**(4): p. 653-664.
100. Çoba, L. and M. Zanker, (2017) "*Replication and Reproduction in Recommender Systems Research-Evidence from a Case-Study with the rrecsys Library".* in *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Springer. pp. 305-314.

101. Li, L., L. Zheng, and T. Li. (2011) "*Logo: a long-short user interest integration in personalized news recommendation"*. in *Proceedings of the fifth ACM conference* on Recommender systems. ACM. pp. 317-320.
102. Du, Y., Wang, J., Wang, X., Chen, J. and Chang, H., (2018) "*Predicting Drug-target Interaction via Wide and Deep Learning"*. in *Proceedings of the 2018 6th International Conference on Bioinformatics and Computational Biology*. ACM. pp. 128-132.
103. Hao, M., S.H. Bryant, and Y. Wang, (2017) "*Predicting drug-target interactions by dual-network integrated logistic matrix factorization"*. Scientific reports, **7**: p. 40376.

**Abstract:**

Recommender systems, which are systems to help users in choosing favourite resources, have been emerged recently, and are being developed rapidly. These systems provide a customized environment via incorporating the past interactions of users and recognizing their interest to choose appropriate resources. Modeling user's behavior and the mechanism of presenting a recommendation are considered as the main issues in developing recommender systems.

Existing methods utilize users' interests as a latent factor in modeling their behavior. The "interest" represents a set of resources with the same characteristics that is chosen by the user in the past. It is inferred from the user's profile. Ignoring the dynamic of a user's interests and missing to prioritize between the different interests in various times are considered as the main challenges in the current methods.

In this research, the "preference", which shows the priority of an interest in a certain time, is considered along with the "interest" to model a user's behavior. The number, time and score of the chosen resources are the main factors in computing a user's preferences. In the proposed method, the behavior of each user is modeled by using the paired $\prec interest, preference \succ$. We utilize the user's profile in a Bayesian network framework named as Dirichlet Process Mixture Model to create user model. Then, similar (neighbor) users are gathered based on the generated behavior model to provide recommendations. Dynamism of the number of interests and adaptive preferences are the main contribution of the proposed method. Also, the proposed model has one hyper parameter which is initialized heuristically.

We used a news and movie recommender system as a case study to evaluate the proposed model. The news dataset was collected from Twitter (news channels). Also, the MovieLens was employed as a movie database. The performance of the proposed model was compared with the state-of-the-art methods such as TimeSVD, Content-based, non-Negative Matrix Factorization, Item-Item Collaborative Filtering. The results of the comparison indicate the efficiency of the proposed model.

Keywords: Recommender System, User Modeling, Hybrid Recommender System, News Recommender System.

Shahrood University of Technology

**Faculty of Computer Engineering**

**Ph.D. Thesis in Artificial Intelligence Engineering**

# News Recommender System Based on User Preferences

By:
Bagher Rahimpour Cami

Supervisor:
Prof. Hamid Hassanpour

Advisor:
Dr. Hoda Mashayekhi

September 2018